

Inhaltsverzeichnis

Übersicht Testkorpus	2
Griechisch	2
Der TLG als Korpusgrundlage	3
1. Bei den Texten des TLG-E handelt es sich um eine Datenbank.	3
2. Eine Datenbank darf zur wissenschaftlichen Forschung vervielfältigt werden.	3
3. Die Texte dürfen umformatiert werden.	3
4. Die Vervielfältigungen dürfen mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufbewahrt werden.	4
5. Die Vervielfältigungen dürfen zur Replikation einzelnen Dritten zur Verfügung gestellt werden	4
6. Die Urheberrechte sind verjährt	4
Latein	4
Die PHI-Datenbank als Korpusgrundlage	5
1. Bei den Texten der PHI-5 handelt es sich um eine Datenbank.	5
2. Eine Datenbank darf zur wissenschaftlichen Forschung vervielfältigt werden.	5
3. Die Texte dürfen umformatiert werden.	6
4. Die Vervielfältigungen dürfen mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufbewahrt werden.	6
5. Die Vervielfältigungen dürfen zur Replikation einzelnen Dritten zur Verfügung gestellt werden	6
6. Die Urheberrechte sind verjährt	6
Abschließende Überlegungen	6

Übersicht Testkorpus

Für die Tests mit den Texten wurden insgesamt vier Anwendungsfälle definiert, jeweils in einer Variante in Griechisch und in Latein.

1. Texte mit nachgewiesener, unzweifelhafter Urheberschaft desselben Ursprungs.
2. Texte mit wissenschaftlich nachgewiesener Pseudo-Autorenschaft in Kombination mit Texten der Autoren, von dem sie angeblich stammen.
3. Texte, die im wissenschaftlichen Diskurs einst als Pseudo-Texte galten und später widerlegt wurden, und deren Urheberschaft wir sicher sein können.
4. Texte von verschiedenen Autoren mit ähnlichen Themen.

Testfall	Griechisch	Latein
1	Plutarch <i>De cohibenda ira</i> Plutarch <i>De tranquillitate animi</i>	Seneca <i>De tranquillitate animi</i> Seneca <i>De ira</i>
2	Plato <i>Phaedo</i> Ps.-Plato <i>Axiochus</i>	Servius und Servius auctus <i>In Vergilii Aeneidos libros</i> <i>In Vergilii Bucolicon librum</i> <i>In Vergilii Georgicon libros</i>
3	Plato <i>Politicus</i> Plato <i>Leges</i>	Cicero <i>Pro Milone</i> Cicero <i>Pro Marcello</i>
4	Plato <i>Apologia Socratis</i> Xenophon <i>Apologia Socratis</i>	Cicero <i>De oratore</i> Tacitus <i>Dialogus de oratoribus</i>

Griechisch

Werk	Abkürzung	TLG-Nr.	Ausgabe des TLG	CTS-URN bei Perseus
Plato <i>Apologia Socratis</i>	Plat. Ap.	0059.002	Burnet, J. Platonis opera, vol. 1, Oxford 1900, St I.17a-42a	urn:cts:greekLit:tlg0059.tlg002
Plato <i>Leges</i>	Plat. Leg.	0059.034	Burnet, J. Platonis opera, vol. 5, Oxford 1907, St II.624a-969d	urn:cts:greekLit:tlg0059.tlg034
Plato <i>Phaedo</i>	Plat. Phd.	0059.004	Burnet, J. Platonis opera, vol. 1, Oxford 1900, St I.57a-118a	urn:cts:greekLit:tlg0059.tlg004
Plato <i>Politicus</i>	Plat. Plt.	0059.008	Burnet, J. Platonis opera, vol. 1, Oxford 1900, St II.257a-311c	urn:cts:greekLit:tlg0059.tlg008
Ps.-Plato <i>Axiochus</i>	Plat. Ax.	0059.038	Burnet, J. Platonis opera, vol. 5, Oxford 1907, St III.364a-372b	urn:cts:greekLit:tlg0059.tlg038
Plutarch <i>De cohibenda ira</i>	Plut. De coh.	0007.095	Pohlenz, M., Plutarchi moralia, vol. 3, Leipzig, 1929, 452f-464d	urn:cts:greekLit:tlg0007.tlg095
Plutarch <i>De tranquillitate animi</i>	Plut. Tranq.	0007.096	Pohlenz, M., Plutarchi moralia, vol. 3, Leipzig, 1929, 464e-477f	urn:cts:greekLit:tlg0007.tlg096

Werk	Abkürzung	TLG-Nr.	Ausgabe des TLG	CTS-URN bei Perseus
Xenophon <i>Apologia Socratis</i>	Xen. Ap.	0032.005	Marchant, E.C., Xenophontis opera omnia, vol. 2, 2nd edn., Oxford: Clarendon Press, 1921 (repr. 1971)	urn:cts:greekLit:tlg0032.tlg005

Der TLG als Korpusgrundlage

Der Thesaurus Linguae Graecae® (TLG®) ist ein 1972 ins Leben gerufene Forschungsprogramm an der Universität von Kalifornien, Irvine, in welchem die meisten literarischen Texte in griechischer Sprache von Homer bis zum Fall von Byzanz 1453 n. Chr. gesammelt und digitalisiert werden. Es existieren mehrere, historisch voneinander abgrenzbare Varianten der Textsammlung:

- TLG Online: <http://stephanus.tlg.uci.edu>. Für die Benutzung ist eine Lizenz obligatorisch, es handelt sich um die aktuelle Textsammlung.
- TLG als CD-Rom: Die Textsammlungen wurden in altgriechischem Betacode per CD-ROM verliehen. Nach Ablauf der Lizenz mussten die CD-ROMs wieder zurückgegeben werden. Für den Gebrauch waren noch zusätzliche Software (z.B. Diogenes: <https://d.iogen.es/>) notwendig. Die letzte Version der auf CD-ROM vertriebenen Sammlung war TLG-E (2000). Davor gab es die Varianten A (1985), C (1988), D (1992).

Für die Nutzung des TLG-E als Korpusgrundlage für die griechischen Werke ergeben sich aus den 2018 mit dem UrhWissG neu eingeführten Bestimmungen nachfolgend skizzierte Szenarien.

1. Bei den Texten des TLG-E handelt es sich um eine Datenbank.

In § 87a UrhG ist eine Begriffsbestimmung „Datenbank“ gegeben. Die per CD-Rom vertriebenen Texte des TLG-E sind nach dieser Definition eine Datenbank, da es sich handelt um:

- eine Sammlung von Werken,
- die systematisch angeordnet sind,
- einzeln mit Hilfe elektronischer Mittel zugänglich sind und
- deren Beschaffung, Überprüfung oder Darstellung eine nach Art oder Umfang wesentliche Investition erfordert.

2. Eine Datenbank darf zur wissenschaftlichen Forschung vervielfältigt werden.

Für die Nutzung von Datenbanken hat der Gesetzgeber Schranken des Rechts des Datenbankherstellers in § 87c UrhG definiert. So ist die Vervielfältigung eines nach Art oder Umfang wesentlichen Teils einer Datenbank zulässig zu Zwecken des Text und Data Mining für Zwecke der wissenschaftlichen Forschung gemäß § 60d UrhG, vgl. auch § 87e UrhG.

3. Die Texte dürfen umformatiert werden.

Bei der Umwandlung in andere Dateiformate handelt es sich nicht um eine Bearbeitung im Sinne des § 23 Absatz 1 UrhG, da der 2018 mit dem UrhWissG neu eingeführte § 23 Absatz 3 UrhG klarstellt, dass die technisch bedingte Änderungen eines Werkes bei Nutzungen für die Fälle des § 60d UrhG

nicht gilt. Mit anderen Worten: Eine Umformatierung ist in diesen Fällen gleichbedeutend mit einer Vervielfältigung.

4. Die Vervielfältigungen dürfen mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufbewahrt werden.

Nach § 60d Absatz 5 UrhG ist es gestattet, die Vervielfältigungen mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufzubewahren, solange sie zur Überprüfung wissenschaftlicher Erkenntnisse erforderlich sind. Eine zeitliche Einschränkung ist nicht gegeben. Da davon auszugehen sein muss, dass die Überprüfung wissenschaftlicher Erkenntnisse im Sinne der Replizierbarkeit von Forschung niemals abgeschlossen ist, müssen, nein dürfen die Vervielfältigungen nicht gelöscht werden nach Abschluss des Forschungsprojektes.

5. Die Vervielfältigungen dürfen zur Replikation einzelnen Dritten zur Verfügung gestellt werden

Einem abgegrenzten Kreis von Personen der gemeinsamen wissenschaftlichen Forschung oder einzelnen Dritten zur Überprüfung der Qualität wissenschaftlicher Forschung dürfen die Vervielfältigungen nach § 60d Absatz 4 UrhG öffentlich zugänglich gemacht werden. Es sind keine zeitlichen Fristen festgelegt, wie lange nach Beendigung eines Forschungsprojektes die Vervielfältigungen einzelnen Dritten öffentlich zugänglich gemacht werden dürfen. Mit anderen Worten: Die dem Korpus zugrunde liegenden Werke dürfen langzeitarchiviert werden, sofern keine unbefugte Benutzung sichergestellt ist.

6. Die Urheberrechte sind verjährt

Gemäß § 87d UrhG sind die Rechte des Datenbankherstellers nach fünfzehn Jahren nach der Veröffentlichung erloschen. Die TLG-E-CD-Rom wurde 2000 veröffentlicht, die Daten darauf sind bis heute unverändert.

Latein

Werk	Abkürzung	PHI-Nr.	Ausgabe der PHI	CTS-URN Perseus	Datierung
Cicero <i>De oratore</i>	Cic. De or.	0474.037	Wilkins, A. S., M. Tulli Ciceronis Rhetorica, vol. 1, 1902	urn:cts:latinLit:phi0474.phi037	55 v. Chr.
Cicero <i>Pro Marcello</i>	Cic. Marcell.	0474.032	Clark, A. C., M. Tulli Ciceronis Orationes, vol. 2, 1918	urn:cts:latinLit:phi0474.phi032	46 v. Chr.
Cicero <i>Pro Milone</i>	Cic. Mil.	0474.031	Clark, A. C., M. Tulli Ciceronis Orationes, vol. 2, 1918	urn:cts:latinLit:phi0474.phi031	52 v. Chr.
Seneca <i>De tranquillitate animi</i>	Sen. Tranq.	1017.012	Reynolds L. D., L. Annaei Senecae Dialogorum, Libri Duodecim, 1977	urn:cts:latinLit:stoa0255.stoa013	unklar (1 - 65 n. Chr.)
Seneca <i>De ira</i>	Sen. De ira	1017.012	Reynolds L. D., L. Annaei Senecae Dialogorum, Libri Duodecim, 1977	urn:cts:latinLit:stoa0255.stoa010	unklar (1 - 65 n. Chr.)

Werk	Abkürzung	PHI-Nr.	Ausgabe der PHI	CTS-URN Perseus	Datierung
Servius und Servius auctus <i>In Vergilii Aeneidos libros</i> <i>In Vergilii Bucolicon librum</i> <i>In Vergilii Georgicon libros</i>	Serv. bzw. Serv. auct. Aen.; Ecl.; Georg.	2349.005; 2349.006; 2349.007	Thilo, G., Servii, Grammatici Qui Feruntur in Vergilii Carmina Commentarii, vols. 1-2, 1878-1884 Thilo, G., Servii; Grammatici Qui Feruntur in Vergilii Bucolica et Georgica; Commentarii, Vol. 3:1, 1887	urn:cts:latinLit:phi2349.phi005 urn:cts:latinLit:phi2349.phi006 urn:cts:latinLit:phi2349.phi007	unklar (spätes 4. - frühes 5. Jahrhundert)
Tacitus <i>Dialogus de oratoribus</i>	Tac. Dial.	1351.003	Furneaux H., Cornelii Taciti Opera, Minora, 1939	urn:cts:latinLit:phi1351.phi003	unklar (ca. 58 - 120 n. Chr.)

Die PHI-Datenbank als Korpusgrundlage

Die PHI-Datenbank ist die vom Packard Humanities Institute zur Verfügung gestellte CD ROM 5.3 (hrsg. 1991), welche im Wesentlichen alle lateinischen literarischen Texte, die vor 200 n. Chr. geschrieben wurden, sowie einige ausgewählte Texte aus der Spätantike enthält. Sie ist im Betacode verfasst und kann z. B. mit [Diogenes](#) ausgelesen werden. Die Texte sind seit 2015 alle auch [online](#) verfü- und durchsuchbar, und zwar nach den ["Fair Use"-Prinzipien](#):

- Zweck und Art der Nutzung, einschließlich der Angabe, ob es sich um eine kommerzielle Nutzung oder um eine Nutzung für nicht gewinnorientierte Bildungszwecke handelt,
- Art des urheberrechtlich geschützten Werks,
- Umfang und Wesentlichkeit des verwendeten Teils im Verhältnis zum urheberrechtlich geschützten Werk als Ganzes,
- Auswirkungen der Nutzung auf den potenziellen Markt für das urheberrechtlich geschützte Werk oder dessen Wert.

Für die Nutzung des PHI-5-Datenbank als Korpusgrundlage für die griechischen Werke ergeben sich aus den 2018 mit dem UrhWissG neu eingeführten Bestimmungen nachfolgend skizzierte Szenarien.

1. Bei den Texten der PHI-5 handelt es sich um eine Datenbank.

In [§ 87a UrhG](#) ist eine Begriffsbestimmung „Datenbank“ gegeben. Die per CD-Rom vertriebenen Texte der PHI-5 sind nach dieser Definition eine Datenbank, da es sich handelt um:

- eine Sammlung von Werken,
- die systematisch angeordnet sind,
- einzeln mit Hilfe elektronischer Mittel zugänglich sind und
- deren Beschaffung, Überprüfung oder Darstellung eine nach Art oder Umfang wesentliche Investition erfordert.

2. Eine Datenbank darf zur wissenschaftlichen Forschung vervielfältigt werden.

Für die Nutzung von Datenbanken hat der Gesetzgeber Schranken des Rechts des Datenbankherstellers in [§ 87c UrhG](#) definiert. So ist die Vervielfältigung eines nach Art oder Umfang wesentlichen Teils einer Datenbank zulässig zu Zwecken des Text und Data Mining für Zwecke der wissenschaftlichen Forschung gemäß [§ 60d UrhG](#), vgl. auch [§ 87e UrhG](#).

3. Die Texte dürfen umformatiert werden.

Bei der Umwandlung in andere Dateiformate handelt es sich nicht um eine Bearbeitung im Sinne des [§ 23 Absatz 1 UrhG](#), da der 2018 mit dem UrhWissG neu eingeführte [§ 23 Absatz 3 UrhG](#) klarstellt, dass die technisch bedingte Änderungen eines Werkes bei Nutzungen für die Fälle des [§ 60d UrhG](#) nicht gilt. Mit anderen Worten: Eine Umformatierung ist in diesen Fällen gleichbedeutend mit einer Vervielfältigung.

4. Die Vervielfältigungen dürfen mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufbewahrt werden.

Nach [§ 60d Absatz 5 UrhG](#) ist es gestattet, die Vervielfältigungen mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufzubewahren, solange sie zur Überprüfung wissenschaftlicher Erkenntnisse erforderlich sind. Eine zeitliche Einschränkung ist nicht gegeben. Da davon auszugehen sein muss, dass die Überprüfung wissenschaftlicher Erkenntnisse im Sinne der Replizierbarkeit von Forschung niemals abgeschlossen ist, müssen, nein dürfen die Vervielfältigungen nicht gelöscht werden nach Abschluss des Forschungsprojektes.

5. Die Vervielfältigungen dürfen zur Replikation einzelnen Dritten zur Verfügung gestellt werden

Einem abgegrenzten Kreis von Personen der gemeinsamen wissenschaftlichen Forschung oder einzelnen Dritten zur Überprüfung der Qualität wissenschaftlicher Forschung dürfen die Vervielfältigungen nach [§ 60d Absatz 4 UrhG](#) öffentlich zugänglich gemacht werden. Es sind keine zeitlichen Fristen festgelegt, wie lange nach Beendigung eines Forschungsprojektes die Vervielfältigungen einzelnen Dritten öffentlich zugänglich gemacht werden dürfen. Mit anderen Worten: Die dem Korpus zugrunde liegenden Werke dürfen langzeitarchiviert werden, sofern keine unbefugte Benutzung sichergestellt ist.

6. Die Urheberrechte sind verjährt

Gemäß [§ 87d UrhG](#) sind die Rechte des Datenbankherstellers nach fünfzehn Jahren nach der Veröffentlichung erloschen. Die PHI-5.3-CD-ROM wurde 1991 veröffentlicht, die 2015 online zur Verfügung gestellte Version ist diejenige von 1991.

Abschließende Überlegungen

Das Testtextcorpus, d.h. auch dessen Herkunft, muss für Nachfolgeforschung transparent gemacht werden. Anderenfalls ist die Replikationsmöglichkeit nicht mehr gegeben. Jegliche willkürliche Veränderungen am Text, um dessen Herkunft zu verschleiern, verbieten sich daher. Veränderungen am Text bzw. in der dazugehörigen XML-Datei sind nur zulässig, wenn sie für die Untersuchung nützlich sind und transparent gemacht werden.

From:

<http://replicatio.science/dokuwiki/> - **documentatio replicationis**

Permanent link:

http://replicatio.science/dokuwiki/doku.php/uebersicht_test_korpus?rev=1791973936

Last update: **2026-10-14**

