

Welche Daten können wir erhalten?

"Preservation Levels" und Erhaltungsmassnahmen im ETH Data Archive

Wenn Daten langfristig verfügbar und nutzbar gehalten werden sollen, ist die Verwendung von offen dokumentierten, weit verbreiteten Standardformaten dringend zu empfehlen. Informationen zur Einschätzung einzelner Formate und Empfehlungen im Hinblick auf die Langzeiterhaltung finden Sie auf der Seite "[Archivtaugliche Dateiformate](#)". Allerdings sind diese Empfehlungen für Forschungsdaten häufig nicht anwendbar. Das ETH Data Archive kann daher grundsätzlich für Daten in beliebigen Dateiformaten genutzt werden. In vielen Fällen kann aber nur eine unveränderte Aufbewahrung gewährleistet werden, während die Nutzbarkeit der Daten in zukünftigen IT-Umgebungen nicht gesichert ist. Dieses anspruchsvolle Ziel kann nur für wenige, gut dokumentierte und offen zugängliche Formate angestrebt werden.

Die Dienstleistungen für den langfristigen Erhalt der im ETH Data Archive aufbewahrten Daten hängen somit nicht von der gewählten Aufbewahrungsfrist, sondern vielmehr von der Eignung des Dateiformats ab.

1. Formatklassifikationen

Nicht geeignet – Bitstream Preservation only

Für Daten mit proprietären oder nicht eindeutig identifizierbaren Formaten (z.B. experimentelle Daten oder Outputs spezifischer Messgeräte) garantieren wir eine unversehrte Speicherung im Originalzustand (Bitstream Preservation). Darüber hinaus können wir aber keine aktiven Erhaltungsmassnahmen leisten. Solche Daten eignen sich denn auch in der Regel nur für eine befristete Aufbewahrung. Dasselbe gilt für verschlüsselte oder anderweitig geschützte Daten, auch wenn deren Formate eigentlich aktive Erhaltungsmassnahmen erlauben würden.

Bedingt geeignet – Limited Preservation Support

Bei einigen Formaten können wir zwar die Entwicklung und mögliche Risiken beobachten und im Einzelfall auch Migrationen durchführen. Doch ist es schwierig, die Folgen einer solchen Transformation abzuschätzen und zu kontrollieren. Dies gilt in der Regel für proprietäre Formate, die weit verbreitet sind (z.B. Microsoft Office Open XML). Die Handlungsmöglichkeiten sind daher eingeschränkt, und vollumfängliche aktive Erhaltungsmassnahmen können nicht garantiert werden.

Die unveränderte Speicherung (Bitstream Preservation) ist auf jeden Fall gewährleistet.

Empfohlen – Full Preservation Support

Auf dieser Ebene werden angemessene Vorkehrungen getroffen, um die Nutzbarkeit der Daten langfristig zu erhalten. Solche aktiven Erhaltungsmassnahmen können nur für Formate geleistet werden, die weit verbreitet und gut dokumentiert sind und offenen Standards folgen.

Wir empfehlen, für die Archivierung von wertvollen, nicht reproduzierbaren Daten wie z.B. langfristigen Beobachtungsdaten und Messreihen nach Möglichkeit ein entsprechendes Standardformat zu wählen.

Die unveränderte Speicherung (Bitstream Preservation) bildet auch hier eine notwendige Voraussetzung für die Erhaltung der Nutzbarkeit.

Für die Zuordnung bestimmter Dateiformate oder Formatgruppen zu diesen drei Klassen siehe unsere [Formatempfehlungen](#). Für Formate, die in dieser Liste nicht erwähnt sind, werden wir keine aktiven Erhaltungsmassnahmen leisten. Die Formatempfehlungen werden regelmässig aktualisiert. Für ein im ETH Data Archive aufbewahrtes Objekt gilt jeweils die aktuelle Klassifizierung, unabhängig von der zum Zeitpunkt des Uploads gültigen Formatempfehlung.

2. Erhaltungsmassnahmen

Massnahme	Full Preservation Support	Limited Preservation Support	Bitstream Preservation only
Formatidentifikation: Eine eindeutige Identifikation des Dateiformats ist unabdingbare Voraussetzung für alle aktiven Massnahmen. Kann ein Format nicht identifiziert werden, so kann nur Bitstream Preservation geleistet werden.	x	x	x

<p>Extraktion und Speicherung von technischen Metadaten:</p> <p>Abhängig von der Verfügbarkeit eines entsprechenden Tools für das jeweilige Format</p>	(x)	(x)	(x)
<p>Validierung gegen etablierte Schemata oder Spezifikationen:</p> <p>Abhängig von der Verfügbarkeit eines entsprechenden Tools für das jeweilige Format</p> <p>Bei Bedarf werden Fehler vor dem Ingest korrigiert.</p>	(x)	(x)	(x)
<p>Formatspezifische Risikoanalyse:</p> <p>Risiken können sich aus dem Dateiformat oder aus bestimmten Eigenschaften in den technischen Metadaten ableiten.</p>	x	x	
<p>Konvertierung vor dem Ingest:</p> <p>Wird im Rahmen definierter Archivierungs-Workflows angeboten und kann (teil-) automatisiert werden. In der Regel werden sowohl das Originalformat als auch das Ergebnis der Konvertierung archiviert.</p>		x	
<p>Periodische Re-Analyse archivierter Daten:</p> <p>Können neue Formate identifiziert werden? Können mittels neuer Extraktoren zusätzliche technische Metadaten analysiert werden?</p>	(x)	x	x
<p>Strategisches Monitoring identifizierter Formate:</p> <p>Hat sich die Risikosituation verändert, droht Obsoleszenz bzw. gibt es ein Nachfolgeformat, das evaluiert werden sollte, um allenfalls eine Konvertierung dorthin vorzunehmen?</p>	x	x	
<p>Migration:</p> <p>Konvertierung des Ausgangsformats in ein neueres Format und Archivierung als neue Version des Objekts. Die frühere Version kann dabei jederzeit rekonstruiert werden.</p>	x	(x)	