

22. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“ am 7. und 8. März 2018 in Marburg

Automatisierte Ingestprozesse in das Transparenzportal Hamburg

Paul Flamme, Staatsarchiv Hamburg

Abstract

Die Bürgerschaft der Freien und Hansestadt Hamburg (FHH) hat im Juni 2012 das Hamburgische Transparenzgesetz (HmbTG) beschlossen und im Oktober 2012 in Kraft treten lassen. Das Gesetz verpflichtet die Behörden und die unter Einfluss der FHH stehenden privaten Unternehmen zur aktiven Veröffentlichung zahlreicher Informationsgegenstände in einem Online-Informationsregister, das kostenlos und anonym in Internet erreichbar sein muss.

Seit Januar 2015 betreibt das Staatsarchiv Hamburg das neu geschaffene Transparenzportal Hamburg. Neben den Stellen der Kernverwaltung Hamburgs (Fachbehörden, Bezirksämter, Landesbetriebe) kommen dort auch 50 Beteiligungsunternehmen ihrer Veröffentlichungspflicht nach. Ende 2017 standen über 80.000 Dokumente und Systeme zum Abruf bereit. Seit dem Start werden im Transparenzportal durchschnittlich circa 1,1 Millionen Seitenaufrufe pro Monat getätigt.

Die Spanne der zu veröffentlichenden Informationsgegenstände beginnt mit Open-Data-Rohdaten wie Statistiken, Geo- und Umweltmessdaten. Sie reicht über Mitteilungen und Beschlüsse des Senats, Gutachten, Verwaltungsvorschriften, Öffentliche Pläne und Baugenehmigungen bis hin zu Subventionen, Zuwendungsvergaben und Verträgen.

Auch wenn es sich beim Transparenzportal der FHH keinesfalls um ein OAIS-konformes Archivsystem handelt gibt es zahlreiche Parallelen zwischen dem Portal und einem digitalen Archiv mit Blick auf Technik und Abläufe.

Die Umsetzung der Veröffentlichungspflicht erfordert ein komplexes technisches System. Bei der IT-Umsetzung konnte nicht auf eine fertige Software „aus dem Regal“ zurückgegriffen werden. Die Gesamtlösung besteht vielmehr aus einer Vielzahl von Teilkomponenten. Dabei wurde soweit möglich auf Open-Source-Software (z.B. die Open Data-Datenkatalogsoftware CKAN, Betriebssystem Ubuntu, Apache-Solr, PostgreSQL) bzw. in der Stadt bereits eingesetzte Verfahren und Softwarelösungen (z.B. das Workflowmanagement-System HIM, SharePoint, Kofax OCR, Redax Schwärzungen, Nachrichtenbroker oder Adobe Reader) zurückgegriffen, welche angepasst bzw. individualisiert wurden.

Über 100 in Frage kommende Liefersysteme waren zu analysieren. Die Daten der veröffentlichungspflichtigen Stellen gelangen über sehr unterschiedliche Wege in das Portal, wobei sich einige Ingest-Grundtypen unterscheiden lassen:

1. Die nicht staatlichen veröffentlichungspflichtigen Stellen erhalten kontrollierte Zugänge ins geschützte Netz der Stadt mit speziellen Tools, die das Hochladen von Daten ins Portal ermöglichen. Die Daten werden im Zentralen Speicher des Systems abgelegt.
2. Veröffentlichungspflichtige Dokumente aus laufenden Registraturen werden über die elektronische Akte mittels eines speziellen teilautomatisierten Workflows in das Informationsregister überwiesen. In Zwischenschritten sind dabei Funktionalitäten wie Formatkonvertierung, das Einfügen einer Texterkennungsschicht, Metadateneingabe, Rechtsprüfung, Schwärzung, Überleitungen in ein Vertragskataster und die Dokumentenfreigabe zu durchlaufen.
3. Der überwiegende Teil der veröffentlichungspflichtigen Datenbestände aus den Fachverfahren der Behörden wird über Schnittstellen automatisiert bereitgestellt. Veröffentlichungs-

pflichtige strukturierte Daten finden sich in über 50 Fachdatenbanken der Hamburger Behörden. Für jedes Liefersystem waren die relevanten Datensätze und die zu veröffentlichenden Feldinhalte zu definieren sowie Lösungen für die jeweiligen Schnittstellen, Formatfestsetzungen, Veröffentlichungsintervalle, Metadatenstruktur etc. zu entwickeln.

Die automatisierte Anbindung reduziert den Aufwand im Betrieb erheblich, weil wiederkehrende manuelle Eingriffe entfallen und Aktualisierungen der Daten automatisch im Portal abgebildet werden. Aus Fachverfahren werden spezielle Dokumente, z.B. Baugenehmigungen, für das Portal generiert und regelmäßig an das Transparenzportal übergeben.

4. Anderer Fachverfahren mit vorhanden Webschnittstellen werden regelmäßig geharvestet, um die veröffentlichungspflichtigen Dokumente für das Transparenzportal zu katalogisieren und zu indiziert. Dabei kommt es vor, dass Metadaten für spezielle Informationsobjekte aus verschiedenen Liefersystemen abgeholt werden müssen. Soweit diese Systeme keine Historisierung oder Versionierung ihrer Daten vorsehen, werden die Primärdaten im Zentralen Speicher des Portals abgelegt.