

Larsen (2012, S. 95) skizziert das digitale Kulturerbe als ein komplexes Phänomen, welchem man sich von verschiedenen Perspektiven (technologisch, sozial, ökonomisch, politisch und/oder benutzungsbezogen) annähern kann. Mason (2007, S. 199-205) geht näher auf den Einfluss, welche die digitale Technologie auf die Sammlung des kulturellen Erbes hat, ein. Denn die sich ständig ändernden Publiziertechiken fordern die sammelnde Institution heraus. Ähnliches drückt auch Jauslin (1995, S. 20) aus, indem er sagt, dass die SNB durch das Sammeln der nationalen Dokumentation viele Veränderungen durchzumachen hat.

Das Internet hat sich zu einem bedeutenden Publikationsmedium entwickelt. Folgerichtig wurde der Sammelauftrag der DNB¹⁹ 2006 um die Archivierung von deutschen Netzpublikationen erweitert (vgl. Erler 2012, S. 15; Sälzer 2012b, S. 37). Dabei hat die NB gemäss Sälzer (2012b, S. 37) Antworten zu folgenden Fragen zu finden: Welche Publikationen im Netz sollen wie gesammelt werden? Wie können sie erschlossen und wie sie zugänglich gemacht und archiviert werden, sodass sie auch in ferner Zukunft noch zugänglich sind?

Unter *Web-Harvesting* versteht man das automatisierte Einsammeln - "Ernten" - von Internet-Dokumenten²⁰, um sie in einem digitalen Archiv einzupflegen und bereitzustellen. Zentrales Element des Web-Harvesting ist eine Software-Komponente, der *Crawler*. Der Crawler sucht ausgehend von einer Liste vorgegebener Webadressen die erreichbaren Dokumente auf und speichert letztere in einer definierten Zielumgebung ab. Beim zielgerichteten Web-Harvesting (*Focused Crawl*) besteht das Ziel darin, möglichst vollständige und konsistente Archivkopien genau derjenigen Websites zu erhalten, deren Adressen in der vorgegebenen Liste enthalten sind. Beim flächigen Web-Harvesting (*Broad Crawl*) wird eine vorgegebene Adressliste lediglich als Einstieg in ein Sammelverfahren verwendet, das sich von Link zu Link weiterarbeitet. Eine Regel dabei könnte lauten, dass zu archivierende Dokumente Bestandteil eines bestimmten Internetbereiches (z.B. Domain '.de') sein müssen, um als archivierungswürdig angesehen zu werden (vgl. Erler 2012, S. 16–17). Das selektive Web-Harvesting geht in die Tiefe, das Domain Crawling mehr in die Breite (vgl. Mason 2007, S. 208).

Hielmcrone (2008, S. 4–5) und Schostag et al. (2012, S. 110) erwähnen, dass in Dänemark folgende drei unterschiedliche Arten von Web-Harvesting angewendet werden, um der gesetzlichen Forderung nach der gänzlichen Abdeckung des dänischen Teils des Webs nachzukommen:

members of an association who receive it as part of their membership available in print and/or online and/or by email") und Website („An information resource suitable for the Internet which is accessible through a web browser. The content is formatted with a markup language and often provides navigation to other web pages via (hypertext) links. Websites are differentiated from online databases by their general lack of internal database management system (DBMS) software although they may have a 'search this site' box (powered by external software like Google, Yahoo!, Ask.com, etc.) that allows a keyword search of the site").

¹⁹ Die Deutsche Bibliothek wird zur DNB mit erweitertem Sammelauftrag inklusive digitale Publikationen (vgl. Die Deutsche Bibliothek 2005, S. 42–43).

²⁰ Anstelle des Ausdruckes *Internet-Dokumente* stehen synonym Begriffe wie *Medienwerke in unkörperlicher Form*, *Netzpublikationen* oder auch *Online-Publikationen*. Darunter werden elektronische Veröffentlichungen verstanden, die über ein öffentliches Netz, heute über das Internet, verfügbar sind (vgl. Matthias et al. 2009, S. 48).