

## Friedrich Kittler's Digital Legacy – PART I - Challenges, Insights and Problem-Solving Approaches in the Editing of Complex Digital Data Collections

Jürgen Enge <juergen\_at\_info-age\_dot\_net>, FHNW Academy of Art and Design, Basel

Heinz WernerKramski <heinz\_dot\_werner\_dot\_kramski\_at\_dla-marbach\_dot\_de>, German Literature Archive Marbach

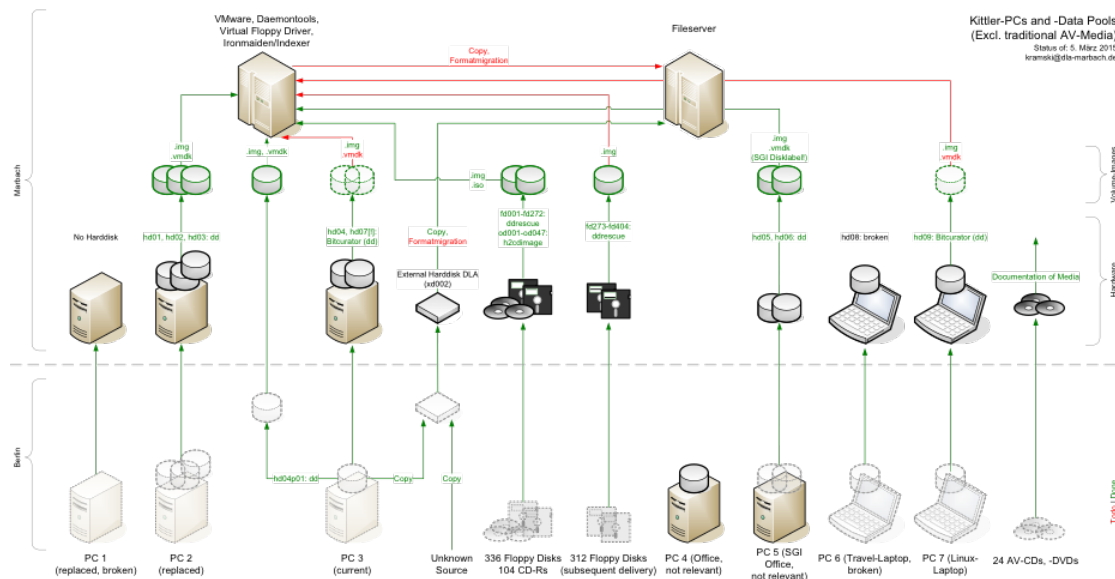
### Abstract

Three years after his death Friedrich Kittler's impact on the Humanities and Media Studies remains a topic of interest to scholars worldwide. The intellectual challenges presented by his theoretical work, however, are now complemented by the practical and archival difficulties of dealing with his personal digital legacy. How are we to preserve, survey and index the complex data collection Kittler bequeathed to the German Literature Archive in Marbach in the shape of old computers and hard drives? How are the Digital Humanities to handle the archive of one of its most important forefathers? To address these questions, this paper will first focus on the estate itself and then describe the design and development of the "Indexer", a tool for the initial indexing of technical information. Two especially problematic aspects are the sheer mass of files (more than 1.5 million) and Kittler's idiosyncratic organization, both of which serve to make conventional content evaluation very difficult. Here, the "Indexer" has proven to be a powerful tool. Finally, a case study using the indexer's web interface will enable us to tackle the question: When and to what purpose did Friedrich Kittler acquire a computer?

### Introduction

Translated by Sandrina Khaled und David Hauptmann

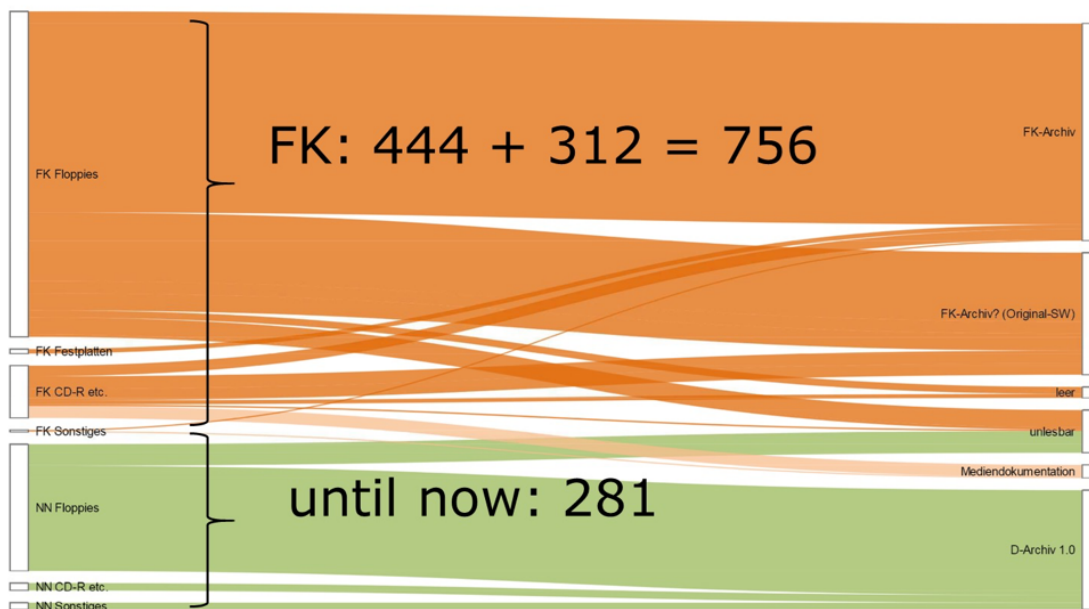
In April 2012 two messy, partially defective, and obviously ancient tower-PCs from the estate of Friedrich Kittler (1943-2011) found their way to the group of Scientific Data Processing of the German Literature Archive Marbach (DLA). A challenging process began: to explore and map Kittler's computers and data storage devices, to identify predecessors and actual systems as well as to separate important from unimportant data. Figure 1 displays the ongoing attempt to obtain an overview of Kittler's PCs and data. With the help of his former collaborators and his heiress Susanne Holl, it soon became obvious that this estate will exceed the DLA's inventory of born-digital items of the last ten years as well as the workflows developed to handle them.



**Figure 1.** Kittler computers and stock of data.

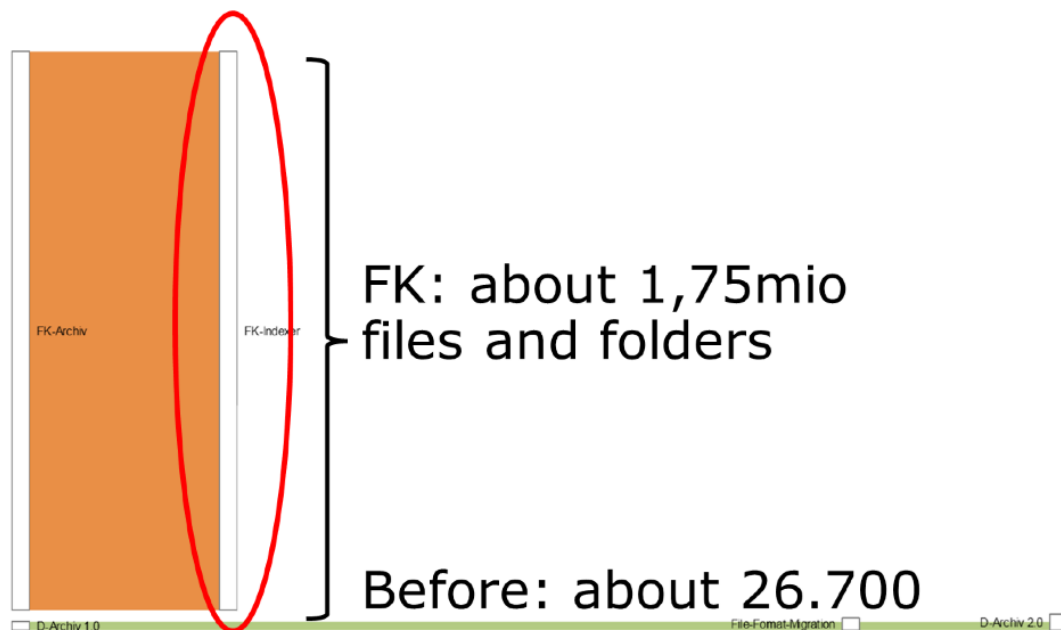
Starting with Thomas Strittmatter's Atari and Macintosh disks in 2003, the DLA developed and established a workflow for preserving and indexing digital personal archives. This worked well with static textual material of manageable size. Until 2013, 281 volumes (mostly floppy disks) from 35 personal archives with 26,700 original files were saved and converted into stable file formats [Bülow 2011]. Almost from the beginning, the DLA adopted a strategy highly recommended by the BitCurator project, that is to file media volumes as a one-to-one copy in sector images [Lee 2013, 27]. In general, sector images are most qualified to preserve technical metadata of file systems (update time stamps, user information etc.) and can moreover be directly integrated as logical drives (read-only) in virtual machines or emulators.

A full-scale comparison of the amount of volumes reveals that Kittler's media (floppies, CD-Rs etc., original hard drives, further non-original transport media and file collections) outnumber the former digital inventory ('D-Archive 1.0') from the last ten years by a significant margin (Figure 2).



**Figure 2.** Number of data volumes scaled according to prior quantity.

The contrast becomes even more apparent in a numerical comparison on the file level (Figure 3). While it is safe to say that most of 1.7 million files are not Kittler's but commercial or free software (ranging from operating system software to special shareware), there is no safe way to automatically detect irrelevant ones and simply leave them apart. This is due to the special complexity of the estate.



**Figure 3.** Number of files scaled according to prior quantity.

Friedrich Kittler bequeathed texts, images, videos, and artifacts of his programming. The latter comprise source code (written in several programming languages) as well as modifications of his hardware configurations. In approximately 30 years this work was executed under different operating systems, starting with MS-DOS, followed by Unix. From approximately 2000 onwards he worked under Gentoo Linux, not as a normal user with limited system rights, but as `root`. His standard working directory was not the common `/home`, he was using `/usr/ich` instead.

The possibility to encounter Kittler's files and source code everywhere made it necessary to show utmost restraint in the exclusion of suggested system directories. Furthermore, it is obvious for Kittler's code that the respective programming environments are of relevance. Therefore, *all* files from hard drives and most of the readable disks and optical media were examined. Only obvious mass products (such as CD-R free samples enclosed in the magazine *c't*) were listed, but not copied. Consequently, a manual selection and conversion as it was carried out in the previous workflow would not work. Not only the amount, but also the specific complexity made evident that this was a special challenge to cope with, from which future estates might benefit.

In July 2012 a group of experts discussed how to approach Kittler's digital estate. Jürgen Enge and Tabea Lurk proposed to examine the untapped data inventory by a full text index for further processing steps. In September 2012 Jürgen Enge presented a prototype working with sample data. In April 2013 substantial partitions and storage devices were analyzed and made available as image copies from which a representative selection was indexed. Since August 2013 "Ironmaiden" (as the package was called in `jest`<sup>[1]</sup>) was consolidated in a virtual machine on the DLA's servers. It was then further developed and eventually confronted with the whole relevant data inventory.

## Preparation Process

Before the actual ingest of data into the Indexer, the preparation process consisted of sector-copying, restoring, and listing. Thereby some practical insights emerged that helped to further develop the workflow at the DLA.

- It was mandatory to label hard disks and removable media before inventory numbers were assigned. We decided on:
  - hd01 etc: hard disks, internal hard drives
  - fd001 etc.: floppy disks
  - od001 etc.: optical disks, CD-ROM, CD-R, CD-RW, DVD etc.
  - xd001 etc.: external files: file collections on further external non-original storage devices, such as DLA's own USB-hard drives used for transport.

These labels also served to create file names for sector images and simplified the administration in internal lists that could later be imported into the Indexer.

- Because there was no useable sequence of volumes available, disks were first of all dated according to the date of the latest file and arranged in a chronological order. Thus disk groups (for instance several disks of a backup set) could be reassembled which was also confirmed by the identical look (color, make, degree of soiling, label, captions). At this stage, a recursive data listing emerged, that could be used for an initial classification of content.
- All reading operations should be carried out under Linux, first of all because many disks were formatted with Linux file systems. Moreover, Linux offers a large range of possible file systems and permits efficient use of the command line in order to automate tasks.
- For sector-wise copying of floppy disks and CD-Rs a new set of scripts and tools was necessary in order to cope with various file systems and deficient media. Here, namely the tools “ddrescue” and “h2cimage” have proven to be very useful. They work well even with deficient media, and a fair amount of data could be restored.<sup>[2]</sup> Further technical problems arose in the attempt to store original files of partitions and optical volumes on the actual file servers of the DLA (as it was previously possible with the floppy disk inventory):
  - Usually, a digital estate is stored on the server file system under an extensive path, named after its holder with systematically labeled sub folders according to the processing state. If additional original files are stored under their own, deeply nested original path, the allowed path length of the operating system might be exceeded.
  - Up-to-date virus scanners often impede the copying of original files contaminated with old (MS-DOS) viruses.
  - DLA's standard file server does not support the original case-sensitive file names (e.g. Makefile vs. makefile) when serving Windows-clients.
  - Reading errors often prevent file-by-file copying of original media.

At this point, preceding its functions as a full text indexing and MIME type analysis tool, the Indexer also serves as a document server which preserves the authentic path information and overcomes the described restrictions of the file systems (for instance through the structuring of its own finely balanced cache).<sup>[3]</sup> Likewise, MS-DOS viruses do not interfere with the analysis under Linux.

10

## Basic components of the Indexer

The Indexer's core is a SOLR-full text index.<sup>[4]</sup> It results from an iterative *identification cascade* and extracts the technologically identified metadata from indexed hard drives and removable media. The full-text index is made accessible through a web-based search and a user interface. It supports the evaluation and decision-making in the archive pre-stage and the subsequent preliminary recording. Before this full-text index is generated, the Indexer creates – using only r/o access – an initial directory for all data in a MySQL-database.<sup>[5]</sup> In addition to a *SessionID*, which through its naming system allows the administration of different file systems/different estates or object groups, the single files are recorded. In doing so, the files technical metadata as well as their immediate surrounding (within the file system) are recorded.<sup>[6]</sup> Now follows an identification cascade that analyzes the data step-by-step and optimizes the identification quality or, respectively, the probability of accurate attribution. Since every tool has its own particular qualities and shortcomings,<sup>[7]</sup> the Indexer combines different tools which can be replaced or upgraded at any time.<sup>[8]</sup>

11

The indexing cascade requires the following tools:

Mandatory are

- Libmagic: creates the initial list of files. It tries to identify MIME type and encoding.
- gvfs-info: has similar capabilities, but can deliver different results.
- sha256sum: creates checksums.
- gzip: compresses extracted full texts for the cache.
- various PHP-extensions.

Highly recommended are

- Apache Tika: extracts not only MIME type and encodings, but also full text in case of texts.
- avconv/ffmpeg: extracts technical metadata from data which gvfs-info identified as time based media (MIME type "video/\*" or "audio/\*")
- ImageMagick: analyzes image- and PDF-data and creates thumbnails for the user interface.

Useful are

- Detex<sup>[9]</sup>: extracts the content (text) from TeX-files (MIME-type "text/x-tex") and removes the TeX-commands.<sup>[10]</sup> Antiword<sup>[11]</sup>: extracts full text from older Word-files (MIME-type "text/application-msword").xsccl.awk<sup>[12]</sup>: extracts comments from source code. The NSRL-library (imported into an Oracle Berkeley DB 4 or 5).
- md5sum: creates a checksum (necessary, when the NSRL-library is used).

The following simplified scheme displays the system architecture:

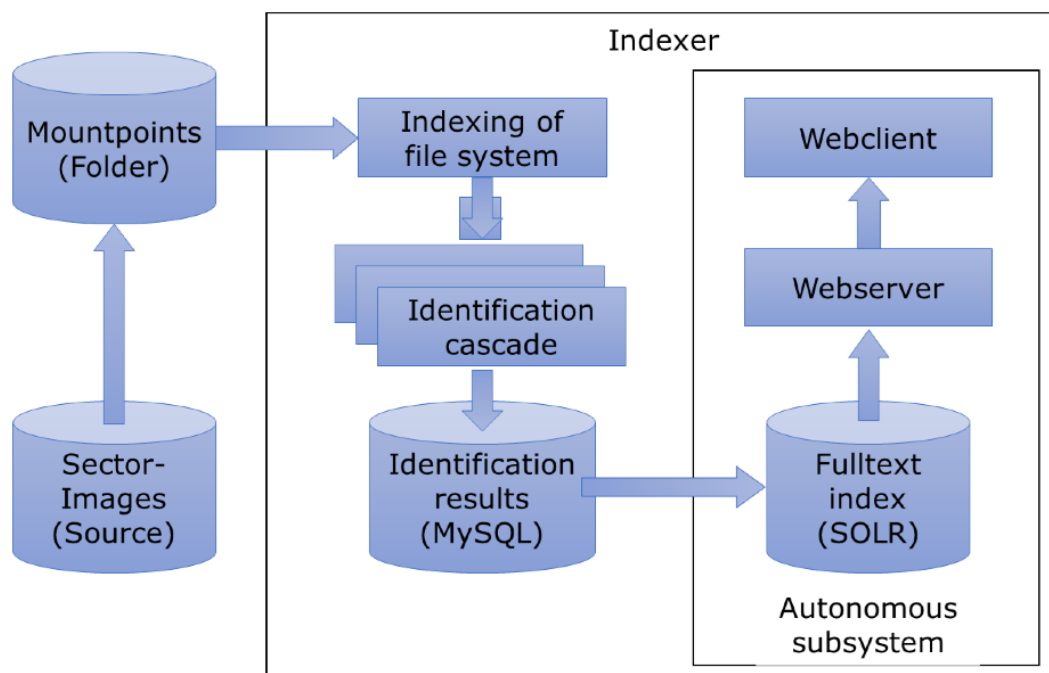


Figure 4. System Architecture.

The Indexer's research and user interface was designed according to the KISS-principle.<sup>[13]</sup> The core functionality of current search engines is created around a simple search slot offering options through a faceted search (see Figure 5) as automatic search suggestions.

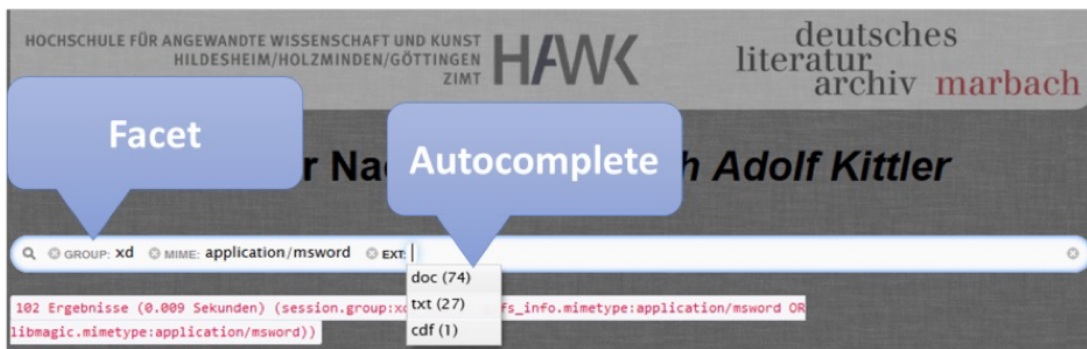


Figure 5. Search field with auto-completion.

The number of results, the search duration (in seconds), and the search query just performed are shown directly under the page header in order to maintain search transparency and comprehensibility. Results are listed below. For every item a preview image, the file's technical metadata, an extract of the full-text in which the search term appears, and a download button, which is active in accordance to the document's clearance level, are displayed.

17

Despite the density of the information, care was taken to keep the display of search results well organized (see Figure 6). Thus, technically versed users could be satisfied, without discouraging non-technical users.

18

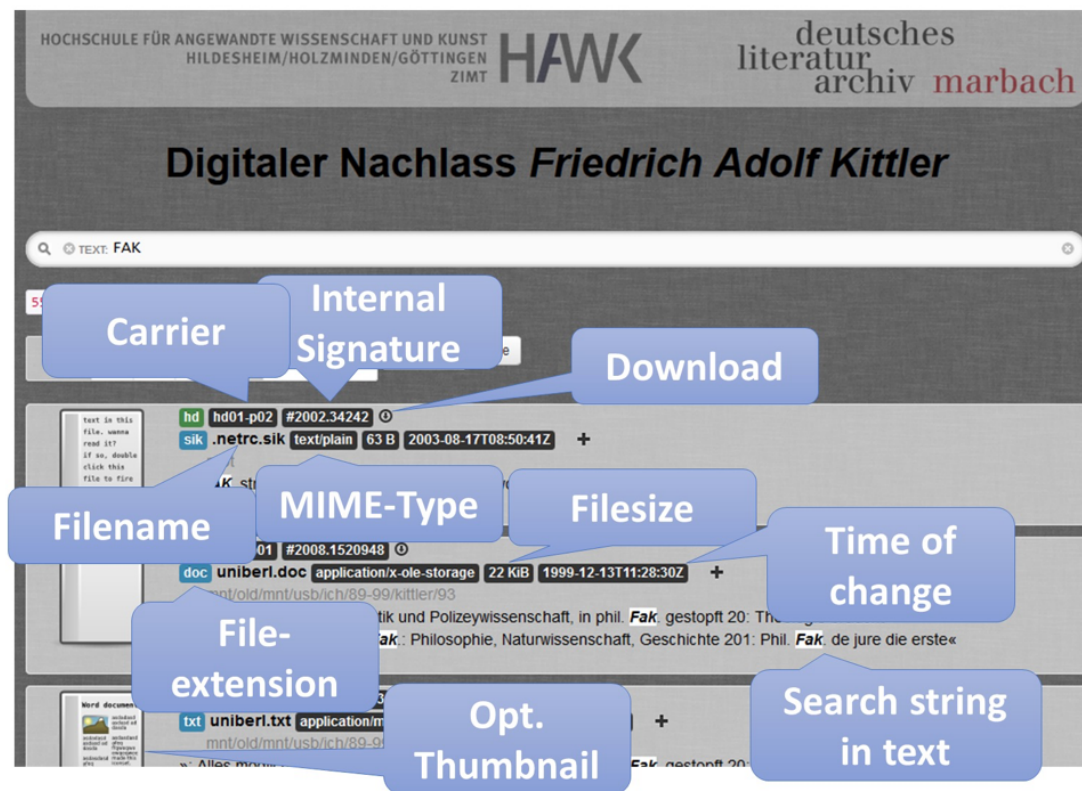


Figure 6. Elements of the search result.

When a record is selected, the partial results of the different analytical tools of the indexing cascade are visible in a subsequent screen. A suggested source quotation to reference the consulted data complements the initial base functionality (see Figure 7 and the citation of Kittler 2011 as an example).

19

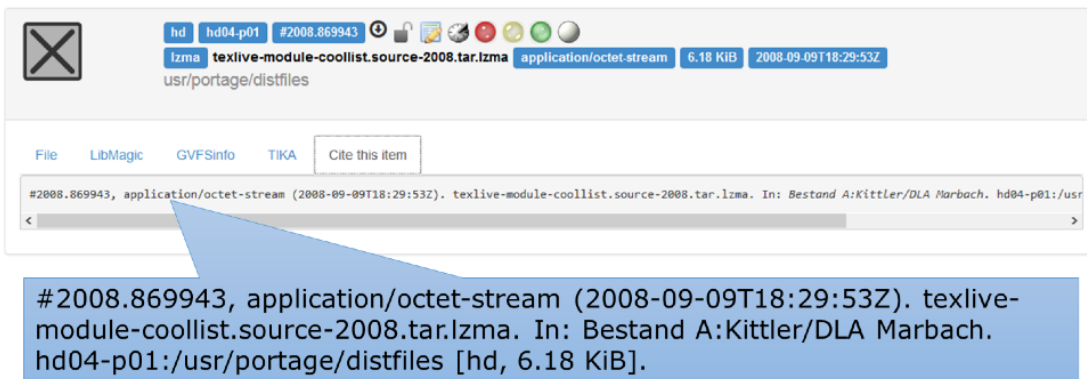


Figure 7. Citation.

## Advancement of the Indexer

The Indexer has been amended in a variety of aspects. As regards to the indexing process, a further module has been integrated in the cascade. As regards to the user interface, an access system and a rudimentary descriptive metadata system have been implemented.

20

## National Software Reference Library

At a talk on the Kittler-project a listener referenced the free “National Software Reference Library (NSRL)” of the American “National Institute of Standards and Technology (NIST)”. Even though the NSRL aims to alleviate forensic analysis in the context of (suspected) cyber crime,<sup>[14]</sup> the basic approach is to eliminate known and non-unique data. This is also useful in the domain of digital archiving [White 2012]. The NSRL version 2.42 from September 2013 of the type *Minimal*, which we integrated into the Indexer for the new recognition module, contains cryptographic hash values (MD5 and SHA-1) of 33,992,326 commercial or free files, as well as information on software packages and manufacturers.<sup>[15]</sup> Figure 8 shows how a Kittler-source code-file CALLTEST.C could be identified through its hash value as part of the software package “Borland C++”. Therefore, it can be identified as an example program and not as source code written by Kittler.

21

sessionid	fileid	name	path	filetype	filesize	md5	filetime	filetime	filetime
4078	1749382	README		file	10579	f4ef68217eac088b85417de0f86d865	1980-01-01 00:00:00	1990-07-06 02:00:00	1980-01-01 00:00:00
4078	1749383	CALLTEST.C		file	389	8c98af7677b8078747a8387d3e49f55	1980-01-01 00:00:00	1990-07-06 02:00:00	1980-01-01 00:00:00
4078	1749384	PLOST.C		file	512	b6c54a9047c021bd21e9a059a8e9d87	1980-01-01 00:00:00	1990-07-06 02:00:00	1980-01-01 00:00:00
4078	1749385	PRIME0.C		file	725	b68dd9aad53be90012614f3937d29b	1980-01-01 00:00:00	1990-07-06 02:00:00	1980-01-01 00:00:00
4078	1749386			file	275	ce4c29a6d5b56e048abb470ca87c44	1980-01-01 00:00:00	1990-07-06 02:00:00	1980-01-01 00:00:00
4078	1749387			file	356	d8954edbe7a07db8644d2acce85cf9	1980-01-01 00:00:00	1990-07-06 02:00:00	1980-01-01 00:00:00

sessionid 4078 =  
Floppy 077, 3.5", vfat,  
ca. 1992

MD5	FileName	FileSize	ProductCode	OpSystemCode	SpecialCode
8CFF8AF7677B8078747A8387D3E49F55	CALLTEST.C	389	5593	Gen	

sessionid	fileid	FileName	File Size	ProductCode	OpSystemcode	Specialcode
4078	1749383	CALLTEST.C	389	5593	Gen	
4078	1749384	PLOST.C	512	5593	Gen	
4078	1749385	PRIME0.C	725	5593	Gen	

ProductCode	ProductName	ProductVersion	OpSystemCode	MfgCode	Language	ApplicationType
5593	Borland C++	Unknown	DOS20	Borland	Unknown	Programming environment
5593	Borland C++	Unknown	Gen	Borland	Unknown	Programming environment
5593	Borland C++	Unknown	WIN30	Borland	Unknown	Programming environment
5593	Borland C++	Unknown	msdos30	Borland	Unknown	Programming environment

Figure 8. Identification of a .c source file with NSRL.

The example also reveals that even in the *minimal* NSRL, the product code is not unique. Due to the NSRL recognition

22



module, we can determine that certain files are not Kittler's. Further identification of software products, especially which system software version Kittler used, cannot be stated precisely. Without any manual research, the tool identified almost 570,000 files (around one third of the total) as user applications or system software mostly irrelevant to further indexing procedures. This is a good hit rate, since Gentoo-Linux is not a binary distribution but has to be fully compiled on its own. It goes without saying that for the most part, the remaining two thirds are also comprised of non-Kittler-files. In close examination, the qualitative gain becomes visible: without any search restrictions, 282,557 files are actually identified as C-source code. If one enters the search term "fak" with which Friedrich Adolf Kittler signed his comments, 1,176 hits remain. Nine of them, however, were found in the NSRL (they use "fak" as the name of variables).

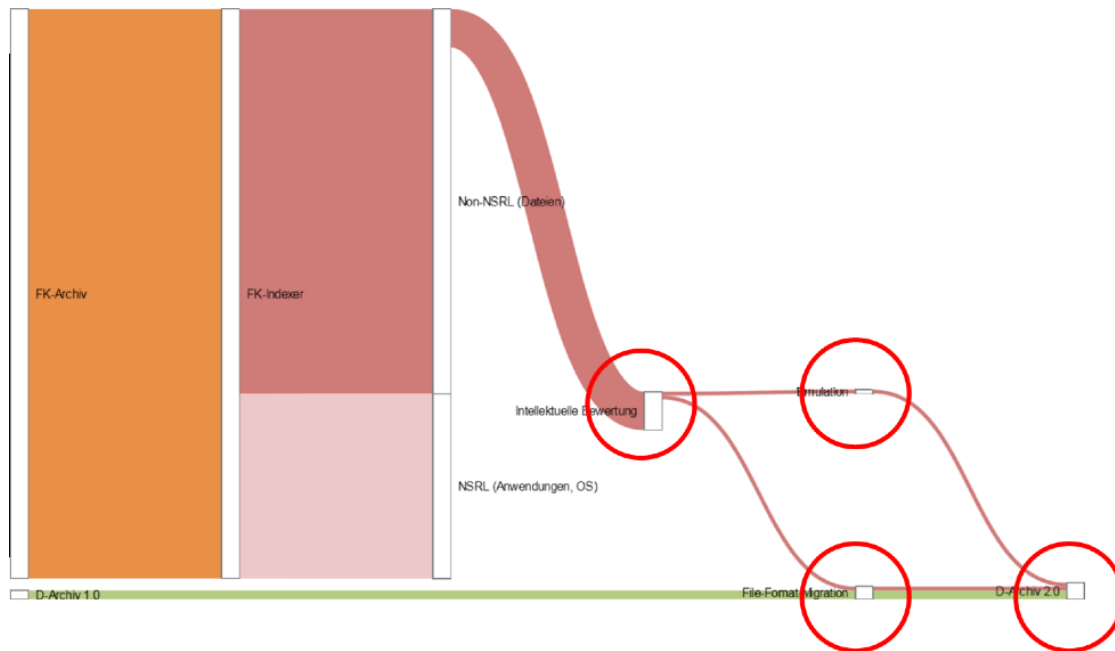


Figure 9. "Distillation" of relevant files.

## Access control

Since the Indexer is used in the archive pre-stage, it is also necessary to assess the recorded content semantically to filter out files, to which access has to be restricted. Even though, for reasons of preservation, no data were deleted, a digital, logical cassation is necessary. This is achieved by a locking and traffic light system regulating access for different user groups.

Currently, there are three roles:

- user (standard)
- editor (scholar, authorized users)
- administrator

Editors and authorized users are allowed to change the "traffic light settings", whereas only administrators can lock or unlock data files for all other roles. This is necessary for private content, which is per agreement locked for sixty years. Locking of data records is in general precarious but absolutely necessary. It is therefore subject to specific security measures and can only be imposed or removed by the administrators or the heiress.<sup>[16]</sup>








	<u>Metadata only,</u>
	<u>Metadata and download (free access),</u>
	<u>Not rated. Review necessary,</u>
	<u>No administrative lock</u>
	<u>Administrative lock</u>

Figure 10. Keys of file states ("lights" and lock)

The system records every change together with the logged-in user names, so that it remains transparent as to who withdrew usage rights from documents. To enable a fast release of more than one document (for instance from a list of search results), dialogue boxes were avoided and the function was directly integrated into the list view. To execute the mass release function/lockage a script in server-side mode is implemented.

26



Figure 11. Search result with "traffic lights" and locking state

## Descriptive metadata

Along with access recommendation, the scholars can classify the entries found with respect to their content, so that the subsequent adoption into the archive can be made more smoothly.

27

Since scholars are not necessarily familiar with the archive's specific classification system, a dialogue box opens up to support the editing of descriptive metadata (see Figure 12). With the dialogue boxes, systematical fast entry (general classification during the archive pre-stage) can be adjusted at any time. A comment field can record important additional information.

28

It is possible in the midterm to provide a discussion forum that allows editors of the archive pre-stage and later on scholars working on the contents to discuss the evaluation. This might be helpful, especially in sketchy and ambiguous cases.

29

Metadaten zu #2001.13535

hd01-p01:SAVE/KOMMENT.SCO

**Ordnungsgattung**

- ☒ Manuskripte
- ☐ Brief-/E-Mail-Wechsel
- ☐ Briefe/E-Mails von
- ☐ Briefe/E-Mails an
- ☐ Dokumente
- ☐ Manuskripte anderer
- ☐ Brief-/E-Mail-Wechsel ohne Bestandsbildner
- ☐ Briefe/E-Mails anderer
- ☐ Dokumente anderer

**Gattung**

- ☐ Bearbeitungen
- ☐ Briefe
- ☐ Briefwechsel
- ☐ Dokumente
- ☐ Dramatisches
- ☐ Gedichte
- ☐ Herausgegebenes
- ☐ Kompositionen
- ☐ Prosa
- ☐ Übersetzungen
- ☒ Verschiedenes
- ☐ Versepen
- ☐ Werke
- ☐ Zeichnungen

**Untergruppe**

- ☐ Abschriften
- ☐ Auswahl
- ☒ Autobiographisches
- ☐ Diktate
- ☐ Gutachten
- ☐ Illustrationen
- ☐ Materialsammlung
- ☐ Mitschriften
- ☐ Reden und Vorträge
- ☐ Rezensionen
- ☐ Sourcecode
- ☐ Verlagsnotizen
- ☐ Vertonungen
- ☐ Verträge
- ☐ Vor- und Nachworte
- ☐ Vorlesungen
- ☐ Widmungen

Kommentar

Technisches Tagebuch (ca. 1991-1992), in dem K. seine Erfahrungen mit SCO-Unix festhält.

Schließen

Speichern

Figure 12. Example dialog for pre-archival indexing.

## The Indexer in practice

The Indexer requires general IT knowledge. The particular components derive from the open-source domain. This allows for the software tool to be available online soon. Because the initial recognition and the identification cascade are CPU-intensive, certain computer benchmarks should be upheld. The following components are necessary for the basic operations of the server system:

- LAMP System as basis for the web component.
- Apache SOLR maintains the full-text index. It should be installed on a sufficiently dimensioned computer system.
- PHP CLI, to execute the command line tools for the indexing.

And the following libraries/frameworks

- HTML/JavaScript
  - Bootstrap<sup>[17]</sup>
  - jquery<sup>[18]</sup>
  - jquery-ui<sup>[19]</sup>
  - backbone.js<sup>[20]</sup>
  - underscore.js<sup>[21]</sup>

30

31

- visualsearch<sup>[22]</sup>
- Backend (PHP)
  - Solarium<sup>[23]</sup>
  - Symfony<sup>[24]</sup>
  - ADOdb<sup>[25]</sup>

The necessary performance capability of the system depends primarily on the amount of data to be indexed. To run a quickly responding web frontend, it is useful to execute the SOLR-index on a server system with sufficient RAM. 32

## Prospect

In summer 2014 two additional old Kittler laptops and 300 5.25" disks emerged which have not yet been processed completely. By now Kittler's latest main computer is physically in Marbach and may contain further material. 33

The Indexer has proven itself as a powerful filtering tool, which automatically classified 30% of Kittler's digital estate. The sophisticated search functions will provide further support for researchers with respect to evaluation and selection, but the amount of required intellectual evaluation is increasing. In several *distillation* steps (whose output is highly speculative), a quantitative proportion will be gained, permitting a targeted file format migration and an emulation of the software components. Since new and valuable tools for ingest and analysis have been successfully developed, the focus is now on questions of presentation and accessibility. The latter implies issues of personal rights (of Kittler and others), which have to be dealt with with utmost care. Text files containing private or professional confidential material (such as letters or academic evaluations, respectively) have to be identified and locked in line with the legal and contractual blocking period. This work has been fairly advanced so the Indexer is now accessible to a broader but still restricted group of researchers, editors, and staff at the DLA. Once this round of intellectual evaluation together with final data protection measures have been completed, the Indexer will be open to public research on-campus at the DLA. Even though, as a kind of "Google for unedited digital estates", Internet access is clearly within technical reach, this would involve further legal consideration since copyrights and personal rights would become an issue. 34

The first impression of 2012 seems to be confirmed. Kittler's digital estate is a testing ground, the complexity of which will not be seen again any time soon. The developed procedures, though, are principally applicable to other digital estates and more hard drives and media from other inheritances have arrived. 35

The Indexer itself, which was developed without any funding, will be continued as a community project, perhaps in cooperation with the BitCurator consortium.<sup>[26]</sup> 36

## Notes

[1] "Intelligent Read-Only Media Identification Engine" or "Intelligent Recursive Online Metadata and Indexing Engine". Due to copyright issues it will be called "Indexer".

[2] The sector-wise copying of disks could not be carried out with the previously used tool "Flopplmg" because a large amount of ext2 and other file systems were included. CD-Rs were converted into .iso-files by the c't-tool "h2cdimage" which creates partially usable images from deficient volumes like ddrescue. In contrast to common copy programs it will not continue reading in deficient sectors without any further progress, so that the drive will not degrade from continued reading attempts (see [Bögeholz 2005]). Kittler comments on one CD-R in the file "komment" (a kind of technical diary): "20.10.10: Viele Dateien auf CD Texte 89-99 sind schon korrupt; arme Nachlaßverwalter, die FAK-Vorlesungsnoten lesen möchten!" ("Many files on the CD [named] Texte 89-99 are already deficient; poor curators who would like to read FAK[Kittler]-lecture notes!") [Kittler 2011a]. It is remarkable to be addressed from the past in such a way. It is also remarkable that with the help of c't – Kittler's favourite computer magazine – an unexpected high amount of backup CD-Rs which Kittler himself expected to be damaged could at least be read [Kittler 2011b].

[3] 273 sector images (exported by an NFS server) were mounted as loop-back devices and indexed.

[4] Information on the SOLR-full text-indexer: <http://lucene.apache.org/solr/>.

[5] Reading access ensures that operations run archive-compliant, so that the system does not leave any traces (own data) within the record. Metadata created within the analysis, as well as the information on the access path is filed beside the record.

[6] The following features were extracted and recorded in a data base: sessionid (ID of the archiving), fileid (distinct file identification number within a session), parentid (ID of the data folder, in which the directory entry is filed), name (file or folder name), path (path of the directory entry), filetype (type, for instance "file", "directory", "reference"), filesize, sha256 (checksum – can be used for authenticity verification), filectime (date of creation), filemtime (changing date), fileatime (date of last access, here often mistakes occur, when data systems are mounted in write mode!), stat (all information of the Unix-call stat( )) archivetime (indexing date). To identify single files a combination of sessionid and fileid is recommended.

[7] The varying results derive from different recognition algorithms and databases within the single tools. Since contradictory statements can occur, the Indexer treats all results as equal, so that the user has to decide which information he should trust.

[8] The following list displays the order of the identification steps.

[9] Information on Detex available at: <https://www.cs.purdue.edu/homes/trinkle/detex/>.

[10] The renunciation of formats optimizes the performance at the full-text search.

[11] Information on Antiword available at: <http://www.winfield.demon.nl/>.

[12] Information on xsccl.awk available at: <http://amit.chakradeo.net/files/xsccl.awk.txt>.

[13] "Keep it simple, stupid", definition online: [https://en.wikipedia.org/wiki/KISS\\_principle](https://en.wikipedia.org/wiki/KISS_principle).

[14] "In most cases, NSRL file data is used to eliminate known files, such as operating system and application files during criminal forensic investigations. This reduces the number of files which must be manually examined and thus increases the efficiency of the investigation."  
[NSRL 2014]

[15] The full version describes 114,095,237 files, but hashes occur repeatedly. Also, the 34 million records of the minimal version were hardly manageable in MySQL, so that for the main table of the NSRL a less demanding Berkeley database was used.

[16] In the case of problematic content where the instance (the file) is still displayed, a disclaimer can be integrated to inform the user, that the content does not derive from the author but is often created by net-related foreign content (spam, ads, junk).

[17] Information on Bootstrap available at: <http://getbootstrap.com/>

[18] Information on jquery available at: <http://jquery.com>

[19] Information on jquery-ui available at: <http://jqueryui.com/>

[20] Information on backbone.js available at: <http://backbonejs.org/>

[21] Information on underscore.js available at: <http://underscorejs.org/>

[22] Information on visualsearch available at: <http://documentcloud.github.io/visualsearch/>

[23] Information on Solarium available at: <http://www.solarium-project.org/>

[24] Information on Symfony available at: <http://symfony.com/>

[25] Information on ADOdb available at: <http://phplens.com/adodb/>

[26] Information on the BitCurator consortium: <http://www.bitcurator.net/>.

## Works Cited

**Bögeholz 2005** Bögeholz, H. "Silberpuzzle. Daten von beschädigten CDs und DVDs retten", 16 (2005): 78–83.  
[http://www.heise.de/artikel-archiv/ct/2005/16/078\\_Silberpuzzle](http://www.heise.de/artikel-archiv/ct/2005/16/078_Silberpuzzle)

**Bülow 2011** Bülow, U. V., Kramski, H. W. "‘Es füllt sich der Speicher mit köstlicher Habe’ – Erfahrungen mit digitalen

Archivmaterialien im Deutschen Literaturarchiv Marbach." In C. Y. Robertson von Trotha, R. Hauser (eds) *Neues Erbe. Aspekte, Perspektiven und Konsequenzen der digitalen Überlieferung*. Karlsruhe (2011): 141–162.  
<http://dx.doi.org/10.5445/KSP/1000024230>

**Kittler 2011a** Kittler, F. *Untitled*. [#1001.10531, text/x-c (2011-08-18T14:37:46Z). *komment*. In: Bestand A: Kittler/DL A Marbach. xd002:/kittler/info [xd, 352.4 KiB].

**Kittler 2011b** Kittler, F. "Wir haben nur uns selber, um daraus zu schöpfen." Interview by Andreas Rosenfelder, *Welt am Sonntag*, January 30 (2011). <http://www.welt.de/print/wams/kultur/article12385926/Wir-haben-nur-uns-selber-um-daraus-zu-schoepfen.html>

**Lee 2013** Lee, C. A., Woods, K., Kirschenbaum, M., Chassanoff, A. *From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions* (2013). <http://www.bitcurator.net/wp-content/uploads/2013/11/From-Bitstream-to-Heritage-S.pdf>

**NSRL 2014** *Introduction to the NSRL* (2014). <http://www.nsrl.nist.gov/new.html>

**White 2012** White, D. R. "The NSRL and Its Potential Role in Digital Curation", Talk at the conference *CurateGear: Enabling the Curation of Digital Collections*, Chapel Hill (2012).  
[http://ils.unc.edu/digccurr/2012Slides/Curategear\\_2012\\_White.pdf](http://ils.unc.edu/digccurr/2012Slides/Curategear_2012_White.pdf)



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.