

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

EMC CORPORATION,
Petitioner,

v.

PERSONALWEB TECHNOLOGIES, LLC and
LEVEL 3 COMMUNICATIONS, LLC,
Patent Owners.

Case IPR2013-00084
Patent 7,945,544 B2

Before KEVIN F. TURNER, JONI Y. CHANG, and
MICHAEL R. ZECHER, *Administrative Patent Judges*.

CHANG, *Administrative Patent Judge*.

FINAL WRITTEN DECISION
35 U.S.C. § 318(a) and 37 C.F.R. § 42.73

I. INTRODUCTION

EMC Corporation (“EMC”) filed a petition on December 16, 2012, requesting an *inter partes* review of claim 1 of U.S. Patent No. 7,945,544 B2 (“the ’544 patent”). Paper 3 (“Pet.”). PersonalWeb Technologies, LLC and Level 3 Communications, LLC (collectively, “PersonalWeb”) filed a patent owner preliminary response. Paper 9 (“Prelim. Resp.”). Taking into account the patent owner preliminary response, the Board determined that the information presented in the petition demonstrated that there was a reasonable likelihood that EMC would prevail with respect to claim 1. Pursuant to 35 U.S.C. § 314, the Board instituted this trial on May 17, 2013, as to claim 1 of the ’544 patent. Paper 14 (“Dec.”).

After institution, PersonalWeb filed a patent owner response (Paper 33 (“PO Resp.”)), and EMC filed a reply to the patent owner response (Paper 40 (“Reply”)). Oral hearing was held on December 16, 2013.¹

We have jurisdiction under 35 U.S.C. § 6(c). This final written decision is entered pursuant to 35 U.S.C. § 318(a). We hold that claim 1 of the ’544 patent is unpatentable under 35 U.S.C. §§ 102 and 103.

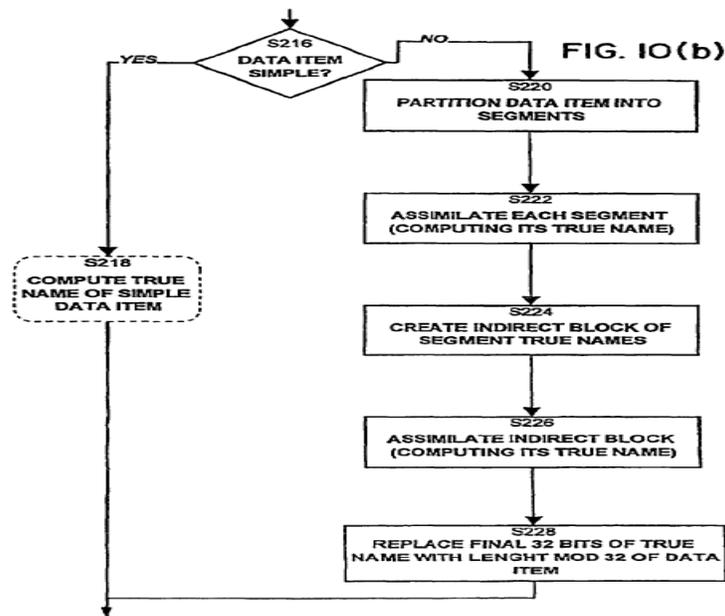
¹ This proceeding, as well as IPR2013-00082, IPR2013-00083, IPR2013-00085, IPR2013-00086, and IPR2013-00087, involve the same parties and similar issues. The oral arguments for all six *inter partes* reviews were merged and conducted at the same time. A transcript of the oral hearing is included in the record as Paper 63.

A. Related Proceeding

EMC indicates that the '544 patent is the subject of litigation titled *PersonalWeb Technologies LLC v. EMC Corporation and VMware, Inc.*, No. 6:11-cv-00660-LED (E.D. Tex.). Pet. 1.

B. The '544 patent

The '544 patent relates to a method for identifying a data item (e.g., a data file or record) in a data processing system, by using an identifier that depends on all of the data in the data item and only on the data in the data item. Ex. 1001, 1:45-49; 3:53-56. Thus, the identity of a data item is said to be independent of its name, origin, location, and address. *Id.* at 3:56-59. According to the '544 patent, it is desirable to have a mechanism for identifying identical data items to reduce duplicate copies of a data item. *Id.* at 3:37-40. Figure 10(b) of the '544 patent, reproduced below, is a flow chart for determining an identifier of a simple or compound data item.



As shown in Figure 10(b) of the '544 patent, for a simple data item (a data item whose size is less than a particular given size) (S216 and S218), a data identifier (True Name) is computed using a function (*e.g.*, a message digest (“MD”) function, such as MD4 or MD5, or a secure hash algorithm (“SHA”) function). *Id.* at 12:18-49, 13:31-42; figs. 10(a) & 10(b). As a result, a data item that has an arbitrary length is reduced to a relatively small, fixed size identifier (True Name) that represents the data item. *Id.*

If the data item is a compound data item (a data item whose size is greater than the particular given size), the system will partition the data item into segments (S220); assimilate each segment (S222); compute the True Name of the segment; create an indirect block consisting of the computed segment True Names (S224); assimilate the indirect block (S226); and replace the final 32 bits of the resulting True Name by the length modulo 32 of the compound data item (S228). *Id.* at 13:43-61, fig. 10(b). The result is the True Name of the compound data item. *Id.*

Figure 11 of the '544 patent is reproduced below:

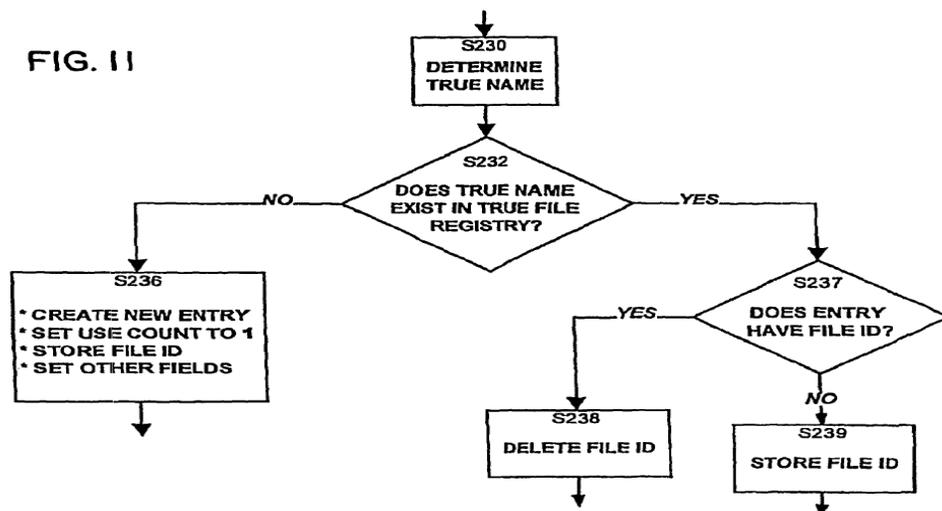


Figure 11 of the '544 patent depicts a mechanism for assimilating a data item into a file system. The purpose of this mechanism is to add a given data item to the True File registry. *Id.* at 14:4-11. If the data item already exists in the registry, the duplicate will be eliminated. *Id.*

To assimilate a data item, the system will determine the True Name of the data item corresponding to the file (S230); look for an entry for the True Name in the True File Registry (S232); and determine whether a True Name entry exists in the True File Registry (S232). *Id.* at 14:4-27, fig. 11. If the entry record includes a corresponding True File ID (Step S237), the system will delete the file (Step S238). Otherwise the system will store the True File ID in the entry record (S239). *Id.* If there is no entry in the True File Registry for the True Name (S232), the system will create a new entry in the True File Registry for the True Name (S236). *Id.*

C. Challenged Claim

According to EMC, claim 1 essentially requires obtaining “values” for two data items, and then comparing these values to ascertain whether the two data items correspond to each other (e.g., whether they are the same).

Pet. 16. Claim 1 recites the following:

1. A computer-implemented method, the method comprising:

(A) for a first data item comprising a first plurality of parts,

(a1) applying a first function to each part of said first plurality of parts to obtain a corresponding part value for each part of said first plurality of parts,

wherein each part of said first plurality of parts comprises a corresponding sequence of bits, and

wherein the part value for each particular part of said first plurality of parts is based, at least in part, on the corresponding bits in the particular part, and

wherein two identical parts will have the same part value as determined using said first function,

wherein said first function comprises a first hash function; and

(a2) *obtaining a first value for the first data item*, said first value obtained by *applying a second function to the part values* of said first plurality of parts of said first data item, said second function comprising a second hash function;

(B) for a second data item comprising a second plurality of parts,

(b1) applying said first function to each part of said second plurality of parts to obtain a corresponding part value for each part of said second plurality of parts,

wherein each part of said second plurality of parts consists of a corresponding sequence of bits, and

wherein the part value for each particular part of said second plurality of parts is based, at least in part, on the corresponding bits in the particular part of the second plurality of parts; and

(b2) *obtaining a second value* for the second data item by applying said second function to the part values of said second plurality of parts of said second data item; and

(C) ascertaining whether or not said *first data item corresponds to said second data item* based, at least in part, on said first value and said second value.

Ex. 1001, 38:34-39:3 (emphases and indentions added).

D. Prior Art Relied Upon

EMC relies upon the following prior art references:

Woodhill US 5,649,196² July 15, 1997 (Ex. 1005)

Frederick W. Kantor, “FWKCS (TM) Contents-Signature System Version 1.22,” FWKCS122.REF (Aug. 10, 1993) (“Kantor,” Ex. 1004)

E. Grounds of Unpatentability

The Board instituted the instant trial based on the following grounds of unpatentability:

| Claim | Basis | References |
|-------|----------|---------------------|
| 1 | § 102(e) | Woodhill |
| 1 | § 102(b) | Kantor |
| 1 | § 103(a) | Kantor and Woodhill |

II. ANALYSIS

A. Claim Construction

We begin our analysis by determining the meaning of the claims. In an *inter partes* review, claim terms in an unexpired patent are given their broadest reasonable construction in light of the specification of the patent in which they appear. 37 C.F.R. § 42.100(b). Under the broadest reasonable construction standard, claim terms are given their ordinary and customary

² Woodhill claims the benefit of U.S. Patent Application No. 08/085,596, which was filed on July 1, 1993.

meaning as would be understood by one of ordinary skill in the art in the context of the entire disclosure. *In re Translogic Tech. Inc.*, 504 F.3d 1249, 1257 (Fed. Cir. 2007). An inventor may rebut that presumption by providing a definition of the term in the specification with reasonable clarity, deliberateness, and precision. *In re Paulsen*, 30 F.3d 1475, 1480 (Fed. Cir. 1994). In the absence of such a definition, limitations are not to be read from the specification into the claims. *In re Van Geuns*, 988 F.2d 1181, 1184 (Fed. Cir. 1993).

In the Decision on Institution, we construed the claim term “data item” to mean “sequence of bits,” and observed that in the context of the specification, the meaning also includes one of the following: (1) the contents of a file; (2) a portion of a file; (3) a page in memory; (4) an object in an object-oriented program; (5) a digital message; (6) a digital scanned image; (7) a part of a video or audio signal; (8) a directory; (9) a record in a database; (10) a location in memory or on a physical device or the like; and (11) any other entity which can be represented by a sequence of bits. Dec. 9. The parties agree with that claim construction. Pet. 6; PO Resp. 1. As noted in the Decision on Institution, that claim construction is consistent with the specification. Dec. 8-9 (citing Ex. 1001, 2:17-18 (“the terms ‘data’ and ‘data item’ as used herein refer to sequences of bits.”); *id.* at 2:18-22, 27-32). We discern no reason to deviate from that claim construction for the purposes of this decision.

B. Principles of Law

To establish anticipation, each and every element in a claim, arranged as recited in the claim, must be found in a single prior art reference. *Net MoneyIN, Inc. v. VeriSign, Inc.*, 545 F.3d 1359, 1369 (Fed. Cir. 2008); *Karsten Mfg. Corp. v. Cleveland Golf Co.*, 242 F.3d 1376, 1383 (Fed. Cir. 2001). We also recognize that prior art references must be “considered together with the knowledge of one of ordinary skill in the pertinent art.” *Paulsen*, 30 F.3d at 1480. Moreover, “it is proper to take into account not only specific teachings of the reference but also the inferences which one skilled in the art would reasonably be expected to draw therefrom.” *In re Preda*, 401 F.2d 825, 826 (CCPA 1968).

A patent claim is unpatentable under 35 U.S.C. § 103(a) if the differences between the claimed subject matter and the prior art are such that the subject matter, as a whole, would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 406 (2007). The question of obviousness is resolved on the basis of underlying factual determinations, including: (1) the scope and content of the prior art; (2) any differences between the claimed subject matter and the prior art; (3) the level of skill in the art; and (4) objective evidence of nonobviousness. *Graham v. John Deere Co. of Kansas City*, 383 U.S. 1, 17-18 (1966). The level of ordinary skill in the art is reflected by the prior art of record. *See Okajima v. Bourdeau*, 261 F.3d 1350, 1355 (Fed. Cir. 2001);

In re GPAC Inc., 57 F.3d 1573, 1579 (Fed. Cir. 1995); *In re Oelrich*, 579 F.2d 86, 91 (CCPA 1978).

We analyze the instituted grounds of unpatentability in accordance with the above-stated principles.

C. Claim 1 – Anticipated by Woodhill

EMC asserts that claim 1 is unpatentable under 35 U.S.C. § 102(e) as anticipated by Woodhill. Pet. 50-57. As support, EMC provides detailed explanations as to how each claim element, arranged as recited in the claim, is disclosed by Woodhill. *Id.* EMC also relies on the declaration of Dr. Douglas W. Clark. Ex. 1009 ¶¶ 43-49.

PersonalWeb counters that Woodhill does not describe all of the limitations of claim 1. PO Resp. 3-15. Specifically, PersonalWeb contends that: (1) Woodhill fails to describe applying a second hash function to shadow files (*id.* at 5-11 (citing Ex. 2016 ¶¶ 25-35)); and (2) Woodhill does not describe binary object identifiers for the first data item and the second data item (*id.* at 11-15 (citing Ex. 2016 ¶¶ 36-40)). PersonalWeb also proffers a declaration of Dr. Robert B. K. Dewar. Ex. 2016 ¶¶ 20-41.

Upon review of the parties' arguments and evidence, we determine that EMC has demonstrated by a preponderance of the evidence that claim 1 is unpatentable under 35 U.S.C. § 102(e) as being anticipated by Woodhill.

Woodhill

Woodhill discloses a system for distributed storage management on a computer network system using binary object identifiers. Ex. 1005, 1:11-17.

The system includes a remote backup file server and a plurality of local area networks in communication with the remote backup file server. *Id.*

Figure 1 of Woodhill, reproduced below, depicts a computer network system that includes a distributed storage management system:

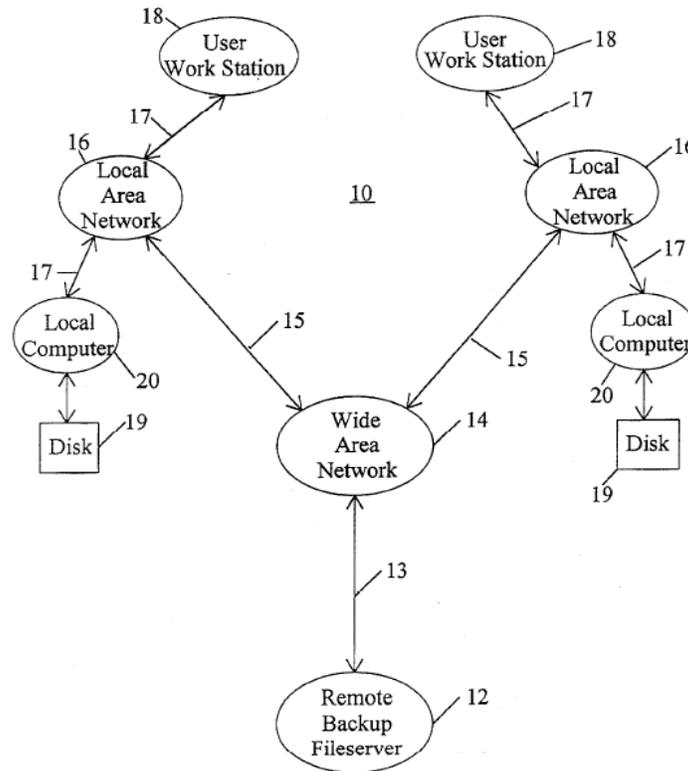


FIG. 1

As illustrated in Figure 1 of Woodhill, remote backup file server 12 communicates with wide area network 14, which communicates with a plurality of local area networks 16. *Id.* at 3:12-30. Each local area network 16 includes multiple user workstations 18 and local computers 20. *Id.* at 3:24-44. The storage space on each disk drive 19 on each local computer 20 is allocated according to the hierarchy illustrated in Figure 2. *Id.* at 3:31-44.

Woodhill's system includes a Distributed Storage Manager (DSM) program for building and maintaining the file database. *Id.* at 3:44-49. The DSM program views a file as a collection of data streams, and divides each data stream into one or more binary objects. *Id.* at 4:13-23; 7:40-43; fig. 5A, item 132. Specifically, data streams represent regular data, extended attribute data, access control list data, etc. *Id.* at 7:44-47. If the size of the data stream is larger than the maximum binary object size, then the DSM program divides the data stream into multiple binary objects; otherwise, a single binary object represents the data stream. *Id.* at 4:23-30; 7:47-59; fig. 5A, items 134 and 136. For each binary object being backed up, a binary object identification record is created in a file database and includes a Binary Object Identifier to identify a particular binary object uniquely. *Id.* at 7:60-8:1; 8:33-34.

Binary object identifiers are calculated based on the contents of the data instead of from an external and arbitrary source so that the binary object identifier changes when the contents of the binary object changes. *Id.* at 8:57-62; 8:40-42. Notably, the binary object identifier includes a binary object hash field that is calculated against the contents of the binary object taken one word (16 bits) at a time using a hash algorithm. *Id.* at 8:22-32. According to Woodhill, duplicate binary objects can be recognized from their identical binary object identifiers, even if the objects reside on different types of computers in a heterogeneous network. *Id.* at 8:62-65.

For large database files on the network computer system, the DSM program utilizes a technique of subdividing the large database files into

granules, and then tracks changes from the previous backup copy of the “granule” level. *Id.* at 14:53-65. This technique is used to reduce the amount of data that must be transmitted to the remote backup file server. *Id.* at 15:4-8. Figure 5G of Woodhill illustrates the “granularization” procedure and is reproduced below:

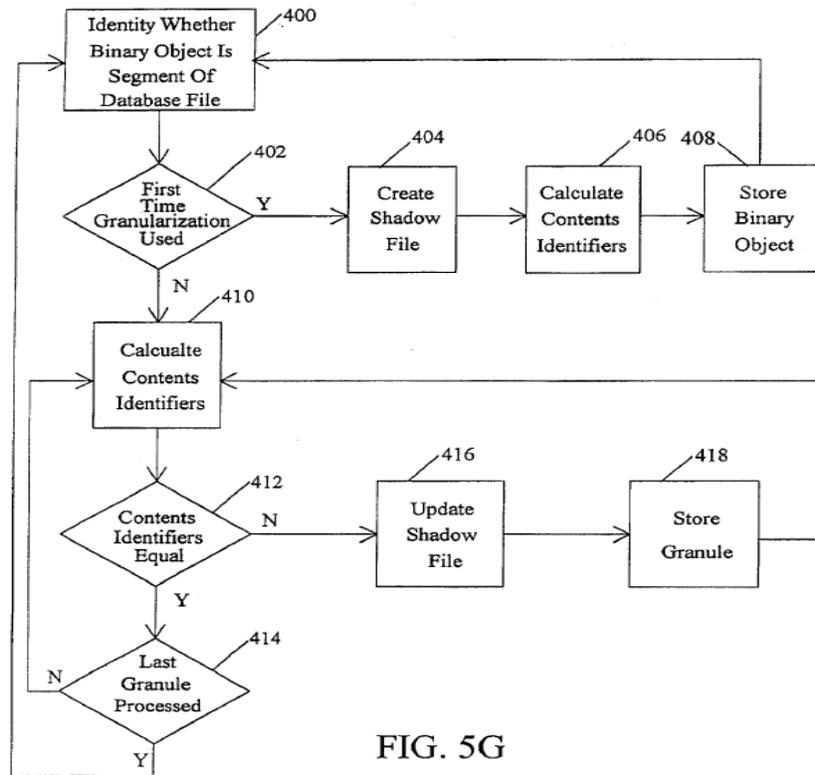


FIG. 5G

As depicted in Figure 5G, if this is the first time that the binary object is being backed up using the “granularization” technique (step 402), the DSM program creates a shadow file, which contains a contents identifier for each granule in the binary object (step 404). *Id.* at 15:9-24. Each contents identifier includes a 32-bit hash number which is calculated against the contents of the granule. *Id.* at 15:24-30; Fig. 5A, step 138.

Each time that the binary object is backed up, the DSM program calculates the contents identifier for each granule in the binary object, and then compares it to the contents identifier of the granule from the last time the binary object was backed up to determine if the granule has changed. *Id.* at 15:32-38. At step 406, the DSM program calculates a change identifier for each granule of the binary object and stores it in the shadow file for that binary object. *Id.* at 15:40-45.

Applying a second hash function to shadow files

Claim 1 requires “obtaining a first value for the first data item, said first value obtained by applying a second [hash] function to the part values of said first plurality of parts of said first data item” (i.e., “a hash of hashes”). In its petition, EMC asserts that Woodhill’s binary object identifiers for the shadow files meet this limitation. Pet. 53-56 (citing Ex. 1005, 5:62-63, 7:60-8:31; 9:6-28; 15:16-24; Ex. 1009 ¶¶ 43-49).

PersonalWeb, however, argues that Woodhill’s granularization process does not disclose applying a second hash function to shadow files. PO Resp. 5, 7 (citing Ex. 2016 ¶¶ 29-35). In particular, PersonalWeb and its expert assert that “Binary Object identifiers 74 are *not* mentioned in connection with Woodhill’s ‘granularization’ procedure, and are not used therein.” PO Resp. 8 (citing Ex. 2016 ¶ 31). PersonalWeb also maintains that EMC’s reliance on Woodhill’s statement that “the default operation is to back up *all files on all disk drives 19 on the local computer 20*” (Ex. 1005, 5:62-63) is incorrect because “Woodhill never describes shadow files as being stored on disk drives 19 of local computers 20.” *Id.* at 10 (citing

Ex. 1005, 15:4-9; Ex. 2016 ¶ 34). Additionally, PersonalWeb, citing to its expert testimony, alleges that a binary object identifier is not created for a shadow file, because the granularization process, in which the shadow files are created, is not used for backing up copies of binary objects for storage on local computers. *Id.* at 8-9 (citing Ex. 2016 ¶¶ 31-35). PersonalWeb further contends that a shadow file will not be backed up by the DMS program, as a shadow file does not meet Woodhill’s definition of a “file” that requires at least two data streams. *Id.* at 11 (citing Ex. 1005, 4:14-15; Ex. 2016 ¶ 35).

In its reply, EMC responds that Woodhill discloses “the application of a hash to the ‘contents identifiers’ in a shadow file.” Reply 1, n.1.

Specifically, EMC alleges that Woodhill discloses calculating a binary object identifier for each shadow file when the DSM program backs up the file. *Id.* at 2. EMC also submits that the shadow file’s binary object identifier is *for* the associated underlying file or binary object. *Id.* at 6-7. We agree with EMC.

PersonalWeb and its expert testimony narrowly focus on Woodhill’s granularization procedure. Notably, Woodhill specifically states that each of the functions performed by the DSM program operates in cooperation with the other functions to form *a unitary computer program*. Ex. 1005, 4:62-5:2; figs. 5a-5l. The disclosure of Woodhill merely divides the DSM program into several distinct functions for explanation purposes. *Id.*

We agree with EMC that Woodhill’s “default operation is to back up all files on all disk drives 19 on the local computer 20” and each *shadow file*, like all files stored on disk drives 19, is divided into one or more binary

objects to be backed up. Pet. 53 (citing Ex. 1005, 5:62-63); 55 (citing Ex. 1009 ¶¶ 46-48; Ex. 1005, 4:13-34; 5:61-63). As noted by EMC, in the process of backing up shadow files, Woodhill would obtain a first value by calculating a binary object identifier (i.e., applying a second hash function) for each shadow file binary object (i.e., the part values – the first hash). Pet. 55-56 (citing Ex. 1009 ¶¶ 45-48; Ex. 1005, 7:60-8:31; 15:16-24).

EMC’s expert, Dr. Clark, testifies:

46. Prior to backing up a binary object using the granularization technique for the first time, the local computer storing the binary object creates a “shadow file” containing the granule contents identifiers for each granule of that binary object. (*Id.* at col. 15, ll. 16-24; Ex 1005.) *Woodhill also discloses claim portions [1c] and [1e]³ through his process of creating shadow files on local computers to store the latest granule contents identifiers for granularized binary objects, and then backup these shadow files.* In particular, a shadow file, including each contents identifier for each granule of a binary object, like any file will be divided into one or more Binary Objects. In some cases, due to the concise nature of a shadow file, a shadow file may be backed up using a single binary object.

47. As I have illustrated, *each shadow file binary object, like all binary objects, has a corresponding Binary Object Identifier.* Further, each Binary Object Identifier includes a hash of the contents of the Binary Object. Consequently, a Binary Object Identifier for a shadow file binary object satisfies these claim elements because it is *a hash (second function) of the*

³ “Claim portions [1c] and [1e]” refer to steps (a2) and (b2) of claim 1. Ex. 1009 ¶ 16.

contents identifiers, or granule hashes (i.e., “part values” of the plurality of parts [granules]).

Ex. 1009 ¶¶ 46-47 (emphases added).

Upon reviewing the evidence on record, we credit the testimony of Dr. Clark over that of Dr. Dewar. *See Yorkey v. Diab*, 601 F.3d 1279, 1284 (Fed. Cir. 2010) (holding that Board has discretion to give more weight to one item of evidence over another “unless no reasonable trier of fact could have done so”). We find that Dr. Clark’s explanations are consistent with Woodhill. *See, e.g.*, Ex. 1005, 4:13-34; 4:62-5:2; 5:61-63; 7:60-8:31; 15:16-24; figs. 5a-5l. On the other hand, Dr. Dewar’s testimony (Ex. 2016 ¶ 34) that shadow files are not stored on the local computers contradicts the disclosure of Woodhill that shadow files are created by the DSM program and stored on the disk drives of the local computers. *See, e.g.*, Ex. 1005, 15:21-24 (The DSM program “creates a ‘shadow file’ which contains a ‘contents identifier’ for each ‘granule’ in the binary object.”); 5:6-9 (The DSM program “operates in the same fashion on each local computer 20 on the network computer system 10.”); 5:7-9; fig. 2, item 24 (The DSM program resides on each disk drive 19 on each local computer 20.); 3:35-49; fig. 3 (The DSM program builds and maintains file database 25, which includes file identification record 34 and binary object identifier 74, on one of disk drives 19 on each local computer 20.); 14:62-65; 15:4-6 (The DSM utilizes the granularization procedure to subdivide large databases files into granules and then tracks changes from the previous backup copy at the

granule level to reduce the amount of data that are being transmitted from the local computer to the remote backup file server.).

To substantiate its position that shadow files are not stored on disk drives 19 on local computers 20, PersonalWeb also relies on Woodhill's statement that the granularization "technique of subdividing files into 'granules' . . . is not utilized in making backup copies of [database file] binary objects for storage on local computers." PO Resp. 10 (citing Ex. 1005, 15:4-9). However, such reliance is misplaced. As EMC notes, reading Woodhill's statement in context, the statement merely confirms that, when backing up large database files using the granularization procedure, the system sends the backup copies of *the database files* to a remote server. Reply 3; *see also* Ex. 1005, 14:59-61 ("As a result, in most cases, the entire 'large' database file would have to be backed up to the remote backup file server 12."). PersonalWeb does not point out where the DSM program would execute the granularization procedure to create the shadow files. Nor does it explain sufficiently why the DSM program would not be executing the granularization procedure *on the local computer*. Given the disclosures of Woodhill noted above, we agree with EMC that the DSM program executes the granularization procedure to create shadow files on disk drive 19 of local computer 20, and not on remote backup file server 12. Reply 3.

We also are not persuaded by PersonalWeb's argument and expert testimony that Woodhill sets forth a definition of the word "file" that requires *at least two data streams*, and that the DMS program would not backup a shadow file to create a binary object identifier, because a shadow

file does not meet that alleged definition of the word “file.” *See* PO Resp. 11; Ex. 2016 ¶¶ 27, 35. PersonalWeb’s argument and expert testimony are not consistent with the explicit disclosure of Woodhill. In particular, they ignore the fact that Woodhill specifically uses the word “file” in the term “shadow *file*.” They also do not provide sufficient explanation why a *shadow file* cannot have more than one data stream or more than one binary object. In fact, a shadow file is consistent with Woodhill’s description of a file. *See* Ex. 1005, 15:21-24 (the DSM program “creates a ‘shadow *file*’ which contains a ‘contents identifier’ for each ‘granule’ in the binary object.”); *id.*, 4:18-19 (“[A] file may contain its *normal data* and may also contain extended *attribute data*.”); *id.*, 2:23-24 (“data files comprised of *one or more* binary objects”) (Emphases added.). As EMC notes, the actual text in Woodhill that PersonalWeb relies on is not a definition of the word “file,” and does not require a file to have *at least two data streams*. Reply 4 (citing Ex. 1005, 4:14-15). Indeed, Woodhill does not preclude a *file* from having *only one* data stream, or *only one* binary object. Ex. 1005, 2:23-24 (“storing data files comprised of *one* or more binary objects”); 4:21-23 (The DMS program “divides each data stream into *one* or more binary objects.”) (Emphasis added.).

For the reasons stated above, EMC has demonstrated by a preponderance of the evidence that Woodhill describes applying a second hash function to shadow files (i.e., “a hash of hashes”).

Shadow file identifiers are for the first and second data items

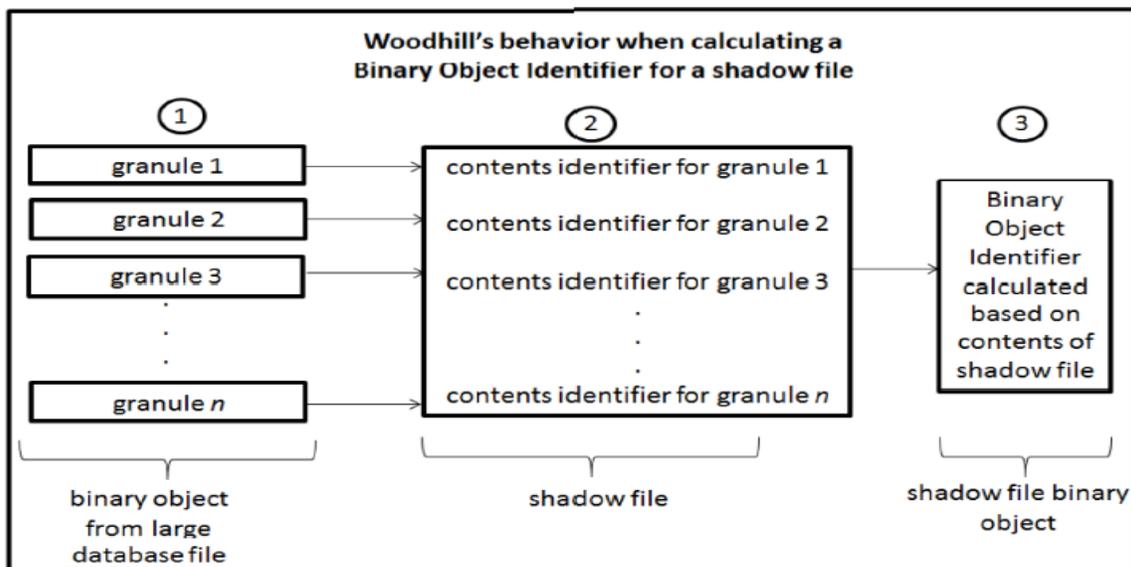
Claim 1 requires “ascertaining whether or not said first data item corresponds to said second data item based, at least in part, on said first value and said second value.” In its petition, EMC takes the position that Woodhill meets this limitation because “by comparing binary objects of successive versions of shadow files, Woodhill by extension compares the binary objects underlying those shadow files.” Pet. 56 (citing Ex. 1005, 9:5-28; Ex. 1009 ¶ 49). EMC further maintains that the comparison is “based, at least in part, on said first value” (the binary object identifier corresponding to a previous version of a shadow file) and “said second value” (the binary object identifier corresponding to the current version of the shadow file). *Id.* at 56-57 (citing Ex. 1005, 9:5-28; Ex. 1009 ¶ 49).

PersonalWeb counters that Woodhill’s shadow file binary object identifiers are not “for the first data item” or “for the second data item.” PO Resp. 11-15 (citing Ex. 2016 ¶¶ 36-40). According to PersonalWeb, “it would be highly unlikely, if not impossible, for a single ‘shadow file’ to be separated from a data stream to form a single standalone ‘binary object,’” and that “the more likely scenario under this assumption would be that a ‘binary object’ would be made up of many shadow files.” *Id.* at 11-12.

In its reply, EMC responds that “PersonalWeb’s assumptions about Woodhill are directly contradictory to Woodhill’s explicit disclosure.” Reply 6 (citing Ex. 1005, 4:13-23; Ex. 1088 ¶¶ 14-15). We agree with EMC. Woodhill expressly discloses *dividing* files into *one or more* data streams, or *one or more* binary objects. Ex. 1005, 2:20-24 (“The present

invention is further directed to a method for the management of storage space . . . storing data files comprised of one or more binary objects.”); 4:22-23 (The DSM program “further divides each data stream into one or more binary objects.”); 4:25-26 (A single binary object may represent a data stream.). Nothing in Woodhill suggests that *a plurality of shadow files* must be *combined* into a *single binary object*.

We also agree with EMC that a binary object identifier for a shadow file is “a hash of hashes” *for the underlying database binary object*. Reply 6-7. As Dr. Clark shows in his illustration (step 1), reproduced below, a binary object for a large database file (a first or second data item) is divided into a plurality of granules (a first or second plurality of parts) (Ex. 1088 ¶¶ 17-18; Ex. 1005, 14:53-15:16):



As shown in step 2 of Dr. Clark’s illustration (Ex. 1088 ¶ 17), Woodhill’s DSM program calculates a contents identifier for each granule of the database binary object, using a hash function (first hash function), and

stores each contents identifier in a shadow file (part value). Ex. 1088 ¶ 20 (citing Ex. 1005, 15:21-28). A binary object identifier (a first or second value) is calculated using a hash function (second hash function) based on the contents of the shadow file (“a hash of hashes”). *Id.* at ¶ 21 (citing Ex. 1005, 8:58-60; 15:21-28); *see also* Ex. 1005, 7:60-8:65. Therefore, the shadow file binary object identifier (step 3) is for the underlying database binary object (step 1). *Id.* We credit Dr. Clark’s testimony as it is consistent with the explicit disclosure of Woodhill.

For the reasons stated above, EMC provides sufficient explanations and evidence to show that Woodhill describes obtaining a first value for the first data item and a second value for the second data item, as well as “ascertaining whether or not said first data item corresponds to said second data item based, at least in part, on said first value and said second value,” as required by claim 1.

Conclusion

For the foregoing reasons, we hold that EMC has demonstrated by a preponderance of the evidence that claim 1 is anticipated by Woodhill.

D. Claim 1 – Anticipated by Kantor

EMC asserts that claim 1 is unpatentable under 35 U.S.C. § 102(b) as anticipated by Kantor. Pet. 28-36. In support of the asserted ground of unpatentability, EMC provides detailed explanations as to how each claim element, arranged as recited in the claim, is disclosed by Kantor. *Id.* EMC

also directs our attention to the declaration of Dr. Clark. *Id.* (citing Ex. 1009 ¶¶ 3-4, 17-25).

In its patent owner response, PersonalWeb counters that Kantor does not describe “applying a first function comprising a hash to each of a plurality of parts of the first data item,” as recited in claim 1. PO Resp. 15-25. PersonalWeb also alleges that Kantor is not a “printed publication” within the meaning of 35 U.S.C. § 102(b). *Id.* at 27-34. In support of its argument, PersonalWeb proffers Dr. Dewar’s declaration (Ex. 2016 ¶¶ 43-55) and Mr. Todd Thompson’s declaration (Ex. 2014).

Upon review of the parties’ arguments and supporting evidence, we determine that EMC has demonstrated by a preponderance of the evidence that claim 1 is unpatentable under 35 U.S.C. § 102(b) as being anticipated by Kantor. We also determine that Kantor is a “printed publication” within the meaning of 35 U.S.C. § 102(b).

Kantor

Kantor describes a method of identifying duplicate files. Ex. 1004, 2-4, 48-49. In particular, Kantor applies a hash function (*e.g.*, a cyclic residue check or cyclic redundancy check (CRC)) to each file within a zip file to obtain the contents signature for each file. *Id.* at 6-8, 48-49. Each contents signature is a string of bits generated from the contents of a file. *Id.*

For each zip file, Kantor creates zip-file contents signatures by hashing the contents signatures for the files contained within the zip file (“a hash of hashes”). *Id.* at 2, 9. As described by Kantor, this is done by “adding together all the 32_bit CRC’s for the files in the zipfile, modulo

2³², separately adding together their uncompressed file_lengths modulo 2³², and then arranging the two resulting hexadecimal numbers as a single structure.” *Id.* at 9. Dr. Clark testifies that addition modulo 2³² is another well-known simple hashing function that uses addition to calculate a value for a file based on the file’s contents. Ex. 1009 ¶ 20. Kantor further compares the zip-file contents signatures to check for duplicate files. Ex. 1004, 2 of Preface, 5, 9.

According to Kantor, contents signatures and zip-file contents signatures are useful to identifying files that have the same contents stored on the electronic bulletin board systems. Ex. 1004, 2 of Preface, 5, 9. For example, when uploading a zip file, the system determines whether that zip file already exists in the system using the zip-file contents signature, and determines whether the inner files of that zip file already exist in the system using the contents signatures for the inner files. *Id.* at 9.

Whether Kantor is a “printed publication”

In its petition, EMC takes the position that Kantor is a “printed publication” under 35 U.S.C. § 102(b). Pet. 28. EMC asserts that Kantor has been publicly available since August 1993, which is prior to the critical date, April 11, 1995, one year before the earliest priority date claimed by the ’544 patent. *Id.* at 3. To substantiate its position, EMC explains that Kantor is “a published manual that describes a software program called the Frederick W. Kantor Contents-Signature System Version 1.22 (‘FWKCS’).” *Id.* at 28 (citing Ex. 1004, Title Page). EMC maintains that Dr. Frederick W. Kantor distributed Kantor—the user manual (version 1.22), the version

relied upon by EMC (*see* Ex. 1004)—with the FWKCS program as shareware and posted it online to electronic Bulletin Board Systems including “The Invention Factory” and “Channel 1” for an extended period of time, where Kantor could be downloaded by anyone. Pet. 3, n.1 (citing Ex. 1004, 3, 158-159). According to EMC, Kantor was accessible to others in the relevant community of the users and system operators of electronic Bulletin Board Systems. *Id.* As support, EMC proffers a declaration of Mr. Michael A. Sussell (Ex. 1049) and declarations of Mr. Jason S. Sadofsky (Exs. 1077, 1087).

In its patent owner response, PersonalWeb counters that Kantor is not a “printed publication.” PO Resp. 27-34. In particular, PersonalWeb alleges that EMC has not established that the specific version of Kantor existed prior to the critical date. *Id.* at 29. PersonalWeb contends that there is no evidence that Kantor was disseminated publicly, catalogued, or indexed in a meaningful way. *Id.* at 32. It is PersonalWeb’s view that EMC fails to establish that one with ordinary skill in the art, exercising reasonable diligence, would have located Kantor prior to the critical date. *Id.* at 30.

We have reviewed parties’ arguments and supporting evidence. Based on the evidence before us, we are not persuaded by PersonalWeb’s arguments. Rather, we determine that EMC has demonstrated by a preponderance of the evidence that Kantor is a “printed publication” within the meaning of 35 U.S.C. § 102(b).

The determination of whether a given reference qualifies as a prior art “printed publication” involves a case-by-case inquiry into the facts and

circumstances surrounding the reference's disclosure to members of the public. *In re Klopfenstein*, 380 F.3d 1345, 1350 (Fed. Cir. 2004). The key inquiry is whether the reference was made "sufficiently accessible to the public interested in the art" before the critical date. *In re Cronyn*, 890 F.2d 1158, 1160 (Fed. Cir. 1989); *In re Wyer*, 655 F.2d 221, 226 (CCPA 1981). "A given reference is 'publicly accessible' upon a satisfactory showing that such document has been disseminated or otherwise made available to the extent that persons interested and ordinarily skilled in the subject matter or art exercising reasonable diligence, can locate it" *Bruckelmyer v. Ground Heaters, Inc.*, 445 F.3d 1374, 1378 (Fed. Cir. 2006) (citation omitted).

Indexing is not "a necessary condition for a reference to be publicly accessible," but is only one among many factors that may bear on public accessibility. *In re Lister*, 583 F.3d 1307, 1312 (Fed. Cir. 2009). In that regard, "while often relevant to public accessibility, evidence of indexing is not an absolute prerequisite to establishing online references . . . as printed publications within the prior art." *Voter Verified, Inc. v. Premier Election Solutions, Inc.*, 698 F.3d 1374, 1380 (Fed. Cir. 2012).

Contrary to PersonalWeb's assertion that Kantor did not exist prior to the critical date and there is no evidence that Kantor was disseminated publicly, Kantor itself shows a copyright date of "1988-1993" and a posted date of "1993 August 10." Ex. 1004, Title Page, the first page after the Title Page ("All of the programs and documents, comprising the entire contents of this Authenticity Verification Zip file FWKCS122.ZIP, together with this

Zipfile itself, are, in accordance with their respective dates of creation or revision, (C) Copyright Frederick W. Kantor 1988-1993.”). Kantor also states:

The FWKCS(TM) Contents_Signature System has become a robust platform for supporting contents_signature functions. FWKCS provides many functions and options for application in a public, commercial, school, institutional, or governmental environment. Extensive technical support is of special value in helping such users to benefit more fully from these many features.

Registered FWKCS hobby BBS users are able to receive a modest amount of assistance, and are invited to participate in the FWKCS conference on The Invention Factory BBS, echoed via Execnet.

Commercial, school, institutional, and governmental users, with their special support needs, are invited to discuss terms for obtaining such assistance.

....

To get a new version of FWKCS, download FWKCSnnn.ZIP from The Invention Factory BBS, where nnn is the new version number without a decimal point. These special downloads are available at no fee, from a 43_line hunt_up group of USR Dual Standard modems, at 2400-16800 bits/sec (including V32.bis).

Ex. 1004, 158-159. It is clear from Kantor that, during the 1988-1993 timeframe, Dr. Kantor had posted many versions of his software and user manual—including Kantor (version 1.22), the version relied upon by EMC (Ex. 1004)—on electronic Bulletin Board Systems.

Mr. Sussell, the co-owner and system operator of the Invention Factory Bulletin Board System, testifies that the Invention Factory Bulletin

Board System is a computer system that allows users to share files, messages, and articles, as well as search, upload, and download files.

Ex. 1049 ¶¶ 3-4. According to Mr. Sussell, he and his wife launched the Invention Factory Bulletin Board System in 1983, and it had over 3,000 subscribers by mid-1993. *Id.* ¶ 6. Mr. Sussell testifies that, by 1993, the system provided all users keyword search functionality and access to various descriptive and meaningful directories. *Id.* ¶¶ 8-10.

More importantly, Mr. Sussell testifies that the Invention Factory Bulletin Board System “extensively utilized and hosted current versions of FWKCS software on its [Bulletin Board System]” and “made publicly accessible and available the complete FWKCS ZIP file that contained both the software as well as related documentation such as user manuals” prior to the critical date. *Id.* ¶ 15; *see also id.* ¶¶ 16-27. Specifically, Mr. Sussell testifies that users would have found Kantor by performing keyword searches on the Invention Factory Bulletin Board System. *Id.* ¶ 21. Mr. Sussell also indicates that the Invention Factory Bulletin Board System advertised Dr. Kantor’s software to its users by including information about the software on the “Welcome” screen, and made the FWKCS Zip file available in four different directories. *Id.* ¶¶ 18-20. Mr. Sussell further testifies that computer disks that contain the FWKCS Zip file were distributed at various Bulletin Board System conferences. *Id.* ¶ 18.

Mr. Sadofsky, a technology archivist and software historian, testifies that he personally verified the authenticity of Kantor—the user manual (version 1.22), the version relied upon by EMC (Ex. 1004)—by comparing it

with a “1993 archived” version, and determined that Kantor is identical to the “1993 archived” version. Ex. 1077 ¶¶ 14-17. Mr. Sadofsky testifies that the source file of the “1993 archived” version has a timestamp of August 10, 1993, at 1:22 AM. *Id.* ¶ 16; *see also* Ex. 1087 ¶¶ 10-11; Ex. 2014 ¶ 5. According to Mr. Sadofsky, Kantor was publicly accessible prior to the critical date. Ex. 1077 ¶¶ 13, 16-17.

PersonalWeb also asserts that Kantor was buried and hidden in the zip file in a manner such that “it would not have been located and accessed by persons interested and ordinarily skilled in the art exercising reasonable diligence even if they had access to the ZIP file.” PO Resp. 33-34 (citing Ex. 2014). However, PersonalWeb’s supporting evidence, Mr. Thompson’s declaration (Ex. 2014), does not substantiate PersonalWeb’s assertion. Upon review of Mr. Thompson’s declaration, we observe that Mr. Thompson downloaded the FWKCS Zip file without any difficulty. Ex. 2014 ¶ 5. Significantly, Mr. Thompson did not follow the instructions provided with the zip file, nor did he use the appropriate computer environment (DOS 3.0 or an IBM OS/2 2.0) that was used normally in 1993-1994 timeframe. Ex. 2014 ¶¶ 6-11; Ex. 1087 ¶¶ 5, 14. Instead, he used non-compatible software (DOS 8.0 and 32-bit Windows XP operating system that was released in 2001). *Id.* Once he followed the instructions and unzipped the FWKCS Zip file, Mr. Thompson located Kantor without difficulty. Ex. 2014 ¶¶ 20-22.

Mr. Sadofsky confirms that the README.TXT file provides simple instructions and, if a user follows the instructions and uses the operating

system that was used normally in 1993-1994 timeframe, the user could locate Kantor without difficulty. Ex. 1087 ¶¶ 13-17. In fact, Mr. Sadofsky demonstrated, in his declaration, several relatively easy ways for a user to access Kantor—with or without installing the software, and with or without help screens. Ex. 1087 ¶¶ 8-16 (II. README.TXT); ¶¶ 17-20 (III. GETLOOK.BAT); ¶¶ 21-22 (IV. FWKCS122 Start Screen and In-Program Help). Based on the evidence before us, we determine that Kantor was available to the extent that persons interested and ordinarily skilled in the art, exercising reasonable diligence, could locate it.

PersonalWeb’s argument that EMC’s witnesses personally did not post or review Kantor prior to the critical date also is unavailing. PO Resp. 29-31 (citing Ex. 2008, 52-55; Ex. 2013, 29-30; Ex. 2015, 98). It is well settled that it is not necessary for the witnesses to have reviewed the reference personally prior to the critical date in order to establish publication. *See In re Hall*, 781 F.2d 897, 899 (Fed. Cir. 1986) (concluding “that competent evidence of the general library practice may be relied upon to establish an approximate time when a thesis became accessible”); *Wyer*, 655 F.2d at 226 (Notwithstanding that there is no evidence concerning actual viewing or dissemination of any copy of the Australian application, the court held that “the contents of the application were sufficiently accessible to the public and to persons skilled in the pertinent art to qualify as a ‘printed publication.’”); *In re Bayer*, 568 F.2d 1357, 1361 (CCPA 1978) (A reference constitutes a “printed publication” under 35 U.S.C. § 102(b) as long as a

presumption is raised that the portion of the public concerned with the art would know of the invention.).

The evidence on this record sufficiently supports that Kantor was posted on a publicly accessible site—the Invention Factory Bulletin Board System—well known to those interested in the art and could be downloaded and retrieved from that site, and therefore Kantor, an electronic publication, is considered a “printed publication” within the meaning of 35 U.S.C. § 102(b). *See Wyer*, 655 F.2d at 226 (An electronic publication, including an on-line database or Internet publication, is considered to be a “printed publication” “upon a satisfactory showing that such document has been disseminated or otherwise made available to the extent that persons interested and ordinarily skilled in the subject matter or art, exercising reasonable diligence, can locate it and recognize and comprehend therefrom the essentials of the claimed invention without need of further research or experimentation.”).

For the foregoing reasons, we determine that EMC has demonstrated by a preponderance of the evidence that Kantor is a “printed publication” within the meaning of 35 U.S.C. § 102(b). Therefore, EMC may rely upon Kantor for its asserted grounds of unpatentability under 35 U.S.C. §§ 102(b) and 103(a).

Applying a hash function to each part of the zip file

Claim 1 requires “for a first data item comprising a first plurality of parts, . . . applying a first function to each part of said first plurality of parts to obtain a corresponding part value for each part of said first plurality of

parts . . . wherein said first function comprises a first hash function.”

Claim 1 also requires “obtaining a first value for the first data item, said first value obtained by applying a second function to the part values of said first plurality of parts of said first data item, said second function comprising a second hash function.”

In its petition, EMC takes the position that Kantor describes the aforementioned limitations. EMC explains that Kantor discloses a “data item” (a zip file) comprising a “first plurality of parts” (the data files within the zip file). Pet. 34 (citing Ex. 1004, 2-3, 48-49; Ex. 1009 ¶ 19). Indeed, Kantor applies a CRC hash function (a first hash function) to the inner files of the zip file (the first plurality of parts) to obtain a contents signature for each inner file (part value). *Id.* (citing Ex. 1004, 48-49; Ex. 1009 ¶ 19). As to the “second hash” limitation, Dr. Clark testifies that Kantor discloses creating zip-file contents signatures for each zip file on the system by hashing the contents signatures for the individual files in the zip file (“a hash of hashes”). Ex. 1009 ¶ 20 (citing Ex. 1004, 9).

In its patent owner response, PersonalWeb counters that Kantor “teaches away” from applying a hash function to each of a plurality of parts of the first data item. PO Resp. 16. PersonalWeb also alleges that Kantor does not apply the CRC hash function to the parts of a zip file because the function is applied to *uncompressed* files before they are *compressed* and packaged into the zip file. *Id.* (citing Ex. 2016 ¶¶ 43-55). According to PersonalWeb, the CRC hash function is applied to different bits (uncompressed files) than the bits (compressed files) that make up the inner

files in the zip file (the alleged data item), and therefore, the CRC hash function is not applied to the *compressed* inner files that are parts of the zip file in determining the zip-file contents signature of the zip file. *Id.* at 16-23.

At the outset, we note that, although a “teaching away” argument could be relevant to an obviousness analysis, “whether a reference teaches away from an invention is inapplicable to an anticipation analysis.”

ClearValue, Inc. v. Pearl River Polymers, Inc., 668 F.3d 1340, 1344 (Fed. Cir. 2012) (citing *Celeritas Techs., Ltd. v. Rockwell Int’l Corp.*, 150 F.3d 1354, 1361 (Fed. Cir. 1998)) (quotation marks omitted).

In any event, we are not persuaded by PersonalWeb’s arguments and expert testimony, as they rest on the erroneous premise that Kantor’s data files contained within a zip file must be *compressed files*. Rather, we agree with EMC that “nothing in Kantor limits the ‘inner files’ of a zip file to *compressed files*” and Kantor’s program works with zip files of all forms. Reply 7-8 (citing Ex. 1088 ¶¶ 26-28; Ex. 1084, 262-63; Ex. 1004, 2, 9). As Dr. Clark notes, PersonalWeb’s evidence shows that zip files are not always compressed, as the standard zip-file format defines seven compression methods, including “Compression method 0,” which does not compress the inner files when packaging them into a zip file. Ex. 1088 ¶ 26 (citing Ex. 2007, 3; Ex. 1084, 262).

Dr. Dewar’s reliance on Kantor’s statements regarding file compression ratio to support his testimony—“Kantor confirms that the inner files in the ZIP files described in Kantor are compressed”—is misplaced. Ex. 2016 ¶ 46 (citing Ex. 1004, 2 of Preface, 9, 55). The mere fact that

Kantor refers to a compression ratio does not support PersonalWeb's position that the inner files of a zip file must be *compressed*, because in the situation where "Compression method 0" is used, which does not compress the inner file, the file compression ratio is one. Contrary to Dr. Dewar's testimony, those portions of Kantor cited by Dr. Dewar do not require the inner files of a zip file to be compressed. Instead, the cited portions of Kantor merely state that the zip-file contents signature *depends on the contents of the files*, and provide examples of items that the zip-file contents signature do not depend upon. Ex. 1004, 2 of Preface ("FWKCS has the special ability to make a 'zipfile contents signature', ('zcs') which is *independent of* . . . the names and dates of files in the zipfile, zipped path information, and file compression ratio."); *id.* at 9 ("This has the desirable property that the resulting *zcs does not depend* on the names of the files, . . . nor on the method nor amount of compression . . ."); *id.* at 55 ("This *zcs does not depend* on the names, dates, compression ratios, order of appearance, zipped paths, nor comments, of files appearing in the zipfile, nor on the zipfile's name, date, nor zipfile comment.") (Emphases added).

We also agree with EMC that, even if Kantor only used compressed inner files, Kantor still would describe the disputed claim limitations, as the first function would *comprise a CRC hash function and a compression function*. Reply 8 (citing Ex. 1088 ¶ 28). Indeed, because claim 1 recites the open-ended phrase "comprising" when describing what the first function includes ("wherein said first function *comprises* a first hash function"), the first function is not limited to just a hash function. PersonalWeb does not

explain adequately why the “first function” cannot comprise more than a hash function. Moreover, *compressing* a file merely changes *the format* of the file, but it does not change *the contents* of the file. In other words, both compressed and uncompressed versions of an inner file have the *same contents* (a corresponding sequence of bits). As discussed above, Kantor’s contents signatures are generated based on *the contents* of the files using the CRC hash function (Ex. 1004, 6-8), and Kantor’s zip-file contents signatures depend on *the contents* of the files and do not depend on the format of the files (Ex. 1004, 2 of Preface, 9, 55). Claim 1 does not place any limitation on *the format* of the plurality of parts (“wherein each part of said first plurality of parts *comprises* a corresponding sequence of bits”). Therefore, PersonalWeb’s argument that the CRC hash function applies to uncompressed files before they are compressed and packaged into the zip file is unavailing.

We are not persuaded by PersonalWeb’s argument that a zip file may include information in addition to the inner files of the zip file (e.g., headers) and, therefore, Kantor’s CRC hash function does not apply to “each part of said first plurality of parts.” That argument is not commensurate within the scope of claim 1. *See In re Self*, 671 F.2d 1344, 1348 (CCPA 1982) (It is well established that limitations not appearing in the claims cannot be relied upon for patentability.). Claim 1 recites “for a first data item *comprising* a first plurality of parts” (emphasis added). As discussed in the claim construction analysis above, the claim term “data item” includes a *portion* of a file. EMC relies on Kantor’s disclosure of a zip file to describe “a first

data item” and the data files within the zip file to describe “a first plurality of parts.” Therefore, we agree with EMC (Reply 10, n.3) that “nothing in claim 1 prohibits the inclusion of [the] addition information in the ‘data item,’ (*comprising* a first plurality of parts).”

Additionally, we are not persuaded by PersonalWeb’s argument that Kantor merely “reads” the CRC values from the zip file and the uncompressed file lengths (sizes) from the zip file, and does not apply a second hash function in determining the zip-file contents signatures. PO Resp. 21. According to Kantor, contents signatures are generated from the contents of the inner files of a zip file by applying a CRC hash function to the inner files. Ex. 1004, 48-49 (“Make a ‘File contents signature’ for (each) File in zipfile(s). . . . The output includes the contents_signature for the file inside the zipfile (using the 32_bit CRC and the uncompressed length of that file). . . .”); *id.* at 1-2 of Preface, 6, 9, 48-49, 55. PersonalWeb also narrowly focuses on Kantor’s reading steps and ignores Kantor’s other steps for determining a zip-file contents signature—“adding together all the 32_bit CRC’s for the files in the zip file, modulo 2^{32} , separately adding together their uncompressed_file_lengths modulo 2^{32} , and then arranging the two resulting hexadecimal numbers as a single structure.” Ex. 1004, 9. Dr. Clark testifies that addition modulo 2^{32} is a well-known simple hashing function that uses addition to calculate a value for a file based on the contents of the file. Ex. 1009 ¶ 20 (citing Ex. 1011). Dr. Clark’s testimony is consistent with Kantor’s disclosure that the resulting zip-file contents signature “does not depend on the names of the files, the dates of the files,

the order in which they appear in the zip file, nor on the method nor amount of compression, nor does it depend on comments.” Ex. 1004, 9.

Given the express disclosure of Kantor, we determine that EMC has demonstrated sufficiently that Kantor’s CRC hash function (a first hash function) applies to “each part of said first plurality of parts” of the first data item, as recited in claim 1.

E. Claim 1 – Obvious over Kantor and Woodhill

EMC asserts that claim 1 is unpatentable under 35 U.S.C. § 103(a) as obvious over Kantor in view of Woodhill. Pet. 36. In particular, EMC submits that “in the event PersonalWeb contends that Kantor does not satisfy the claim limitation of a ‘plurality of parts’ of a data item, a person of ordinary skill would have found it obvious to modify Kantor to meet that limitation.” *Id.* EMC maintains that dividing a file into parts (e.g., dividing a file into a plurality of binary objects or granules) was a well-known technique to handle large files, as evidenced by Woodhill, to reduce the amount of data that must be transmitted. *Id.* (citing Ex. 1005, 4:14-30; 14:52-15:8; Ex. 1009 ¶ 26).

In its patent owner response, PersonalWeb counters that the obviousness ground of unpatentability does not cure the deficiencies of Kantor. PO Resp. 25. PersonalWeb essentially relies upon the same arguments presented above with respect to the anticipation ground of unpatentability based on Kantor. *Id.* at 25-26. As discussed above, we have addressed those arguments and determined that they are unavailing.

PersonalWeb also alleges that one with ordinary skill in the art would not have modified Kantor to include small data items, because “there is no need for this in Kantor as Kantor is concerned with avoiding duplicate files and not with creating duplicates by backing up files.” *Id.* at 26. However, PersonalWeb’s argument improperly focuses on Woodhill’s *entire back-up procedure*. EMC’s proposed modification does not require incorporating Woodhill’s *entire back-up procedure* into Kantor’s method of identifying duplicate files. In fact, EMC merely relies upon Woodhill’s technique of *dividing a file into a plurality of parts*. Pet. 36. “It is well-established that a determination of obviousness based on teachings from multiple references does not require an actual, physical substitution of elements.” *In re Mouttet*, 686 F.3d 1322, 1332 (Fed. Cir. 2012); *see also In re Etter*, 756 F.2d 852, 859 (Fed. Cir. 1985) (en banc) (“[T]he criterion [is] not whether the references could be physically combined but whether the claimed inventions are rendered obvious by the prior art as a whole.”). “To justify combining reference teachings in support of a [ground of unpatentability] it is not necessary that a device shown in one reference can be physically inserted into the device of the other.” *In re Keller*, 642 F.2d 413, 425 (CCPA 1981).

Moreover, incorporating Woodhill’s technique of dividing a file into a plurality of parts (Ex. 1005, 4:14-30; 14:52-15:8; Ex. 1009 ¶ 26) into Kantor’s method of identifying duplicate files would not have been beyond the level of an ordinarily skilled artisan. *KSR*, 550 U.S. at 417 (“[I]f a technique has been used to improve one device, and a person of ordinary skill in the art would recognize that it would improve similar devices in the

same way, using the technique is obvious unless its actual application is beyond his or her skill.”).

PersonalWeb further submits that its evidence of non-obviousness outweighs EMC’s evidence of obviousness. PO Resp. 26-27. In support of its argument, PersonalWeb directs our attention to three licensing agreements, as well as the declaration of Mr. Kevin Bermeister. *Id.* at 27 (citing Exs. 2010-12; Ex. 2009 ¶¶ 3-9). PersonalWeb argues that each license granted to a third party was not for the purpose of settling a patent infringement suit. *Id.*

In its Reply, EMC contends that PersonalWeb has failed to establish a sufficient nexus between claim 1 of the ’544 patent and the above-identified license agreements. Reply 12-13. EMC argues that each of the licenses granted rights to more than just claim 1, and involved related parties with interlocking ownership and business interests. *Id.* We agree with EMC that PersonalWeb has failed to establish the requisite nexus between the licensing agreements and claim 1.

A party relying on licensing activities as evidence of non-obviousness must demonstrate a nexus between those activities and the subject matter of the claims at issue. *GPAC*, 57 F.3d at 1580. Further, without a showing of nexus, “the mere existence of . . . licenses is insufficient to overcome the conclusion of obviousness” when there is a strong ground of unpatentability based on obviousness. *SIBIA Neurosciences, Inc. v. Cadus Pharm. Corp.*, 225 F.3d 1349, 1358 (Fed. Cir. 2000); *Iron Grip Barbell Co. v. USA Sports, Inc.*, 392 F.3d 1317, 1324 (Fed. Cir. 2004).

The evidence of non-obviousness presented by PersonalWeb falls short of demonstrating the required nexus. Neither PersonalWeb nor the declaration of Mr. Bermeister (Ex. 2009) establishes that the licensing agreements (Exs. 2010, 2011, 2012) are directed to the claimed subject matter recited in claim 1. For instance, PersonalWeb does not present credible or sufficient evidence that the three licensing agreements arose out of recognition and acceptance of the claimed subject matter recited in claim 1. In the absence of an established nexus with the claimed invention, secondary consideration factors are entitled little weight, and generally have no bearing on the legal issue of obviousness. *See In re Vamco Machine & Tool, Inc.*, 752 F.2d 1564, 1577 (Fed. Cir. 1985). Furthermore, even if we assume that above-identified licenses establish some degree of industry respect for the claimed subject matter recited in claim 1, that success is outweighed by the strong evidence of obviousness over Kantor and Woodhill discussed above.

Based on the record before us, including the evidence of non-obviousness presented by PersonalWeb and the evidence of obviousness presented by EMC, we conclude that EMC has demonstrated by a preponderance of the evidence that claim 1 would have been obvious over the combination of Kantor and Woodhill.

F. EMC's Motion to Exclude

EMC seeks to exclude the following exhibits: (1) three license agreements (Exs. 2010-12); (2) Mr. Bermeister's declarations (Exs. 2009, 2017) relating to those license agreements; and (3) Mr. Thompson's

declaration (Ex. 2014). Paper 50 (“Pet. Mot.”). PersonalWeb filed the license agreements and Mr. Bermeister’s declarations as evidence of non-obviousness to rebut EMC’s assertion that claim 1 would have been obvious over the combination of Kantor and Woodhill. PO Resp. 12-13. As to Mr. Thompson’s declaration, PersonalWeb proffered that evidence to support its assertion that Kantor—a user manual that was disseminated publicly with the software in a zip file—was not made sufficiently accessible to a person interested and ordinarily skilled in the art. *Id.* at 32-34.

With respect to the license agreements and Mr. Bermeister’s declarations (Exs. 2010-2012; Exs. 2009, 2017), EMC argues that they are irrelevant under Federal Rule of Evidence 402⁴, highly prejudicial, confusing, and misleading under Federal Rule of Evidence 403. Pet. Mot. 1-13. As to Mr. Thompson’s declaration, EMC argues that it should be excluded under Federal Rule of Evidence 402. *Id.* at 14-15. In particular, EMC alleges that: (1) Mr. Thompson does not possess the skill of a person within ordinary skill in the art (*id.* at 14, citing Ex. 1082, 13-14); (2) Mr. Thompson did not use compatible software from the relevant time period (*id.* at 14, citing Ex. 1082, 40-41; Ex. 2014, 4, 6); and (3) Mr. Thompson did not follow the instructions provided with the zip file (*id.* at 14, citing Ex. 1082, 32-35).

The current situation does not require us to assess the merits of EMC’s motion to exclude. As discussed above, even without excluding

⁴ As stated in 37 C.F.R. § 42.62, the Federal Rules of Evidence generally apply to proceedings, including *inter partes* reviews.

PersonalWeb's supporting evidence, we have determined that Kantor is a "printed publication" under 35 U.S.C. § 102(b), and EMC has demonstrated by a preponderance of the evidence that claim 1 is unpatentable over the combination of Kantor and Woodhill.

Accordingly, EMC's motion to exclude evidence is *dismissed* as moot.

G. PersonalWeb's Motion to Exclude

PersonalWeb seeks to exclude the following items of evidence:

(1) Kantor (Ex. 1004); (2) certain documents that corroborating witnesses' knowledge and recollections (Exs. 1046-1048, 1051-1054, 1073, 1074, 1079-1081) and the portions of witnesses' testimony regarding these documents; (3) the declarations of Messrs. Sussell and Sadofsky (Exs. 1049, 1077, 1087) and Mr. Sadofsky's deposition (Ex. 2013, 30, 66); and (4) Dr. Clark's rebuttal declaration (Ex. 1088 ¶¶ 26-27, 30). Paper 48 ("PO Mot.").

EMC opposes PersonalWeb's motion to exclude. Paper 55 ("Opp."). In response, PersonalWeb filed a reply to EMC's opposition to its motion to exclude. Paper 58 ("PO Reply"). For the reasons stated below, PersonalWeb's motion to exclude is *denied*.

1. Kantor

PersonalWeb alleges that Kantor should be excluded as unauthenticated and inadmissible hearsay under Federal Rules of Evidence 901 and 902. PO Mot. 1, 6. In particular, PersonalWeb argues that "[n]o witness of record has personal knowledge of Kantor existing prior to [the

critical date], and electronic data such as Kantor is inherently untrustworthy because it can be manipulated from virtually any location at any time.” *Id.* at 2-4. According to PersonalWeb, the dates provided by Kantor are inadmissible hearsay because Kantor is not self-authenticating. *Id.* at 2, 5-6.

EMC argues that Kantor has been authenticated under Federal Rules of Evidence 901, and that the document is not hearsay, because it is being offered for what it describes—not for the truth of its disclosures. *Opp.* 1-8. In particular, EMC disagrees with PersonalWeb that Kantor cannot be authenticated without direct testimony from a witness with personal knowledge that Kantor existed prior to the critical date. *Opp.* 1. EMC asserts that it need “only produce evidence ‘sufficient to support a finding’ that the reference ‘is what the proponent claims it is.’” *Id.* at 1-2 (citing Fed. R. Evid. 901(a)). EMC also contends that testimony from Messrs. Sussell and Sadofsky provides sufficient evidence to authenticate Kantor. *Opp.* 1-6 (citing Exs. 1049, 1077, 1087).

In its reply, PersonalWeb argues that Federal Rules of Evidence identified by EMC are not applicable to Kantor, because Mr. Sussell did not post or review Kantor prior to critical date. *PO Reply* 1-5 (citing Ex. 2008, 52-55, 65). PersonalWeb also alleges that Kantor’s authenticity is suspicious, as electronic data are inherently untrustworthy and there is no chain of custody. *Id.*

We have considered PersonalWeb’s arguments as well as EMC’s contentions and supporting evidence. We are not persuaded that Kantor should be excluded. At the outset, we disagree with PersonalWeb’s position

that a witness cannot authenticate a document, unless the witness is the author of the document or the witness has reviewed the document prior to the critical date. Federal Rule of Evidence 901(a) states that the authentication requirement is satisfied if the proponent presents “evidence sufficient to support a finding that the item is what the proponent claims it is.” Therefore, neither a declaration from the author, nor evidence of someone actually viewing the document *prior to critical date*, is required to support a finding that the document is what it claims to be. *See Hall*, 781 F.2d at 899 (concluding “that competent evidence of the general library practice may be relied upon to establish an approximate time when a thesis became accessible.”); *Wyer*, 655 F.2d at 226 (Notwithstanding that there is no evidence concerning actual viewing or dissemination of any copy of the Australian application, the court held that “the contents of the application were sufficiently accessible to the public and to persons skilled in the pertinent art to qualify as a ‘printed publication.’”).

Further, it is well settled that an uninterrupted chain of custody is not a prerequisite to admissibility, but rather gaps in the chain go to weight of the evidence. *U.S. v. Wheeler*, 800 F.2d 100, 106 (7th Cir. 1986); *see also U.S. v. Aviles*, 623 F.2d 1192, 1198 (7th Cir. 1980) (“If the trial judge is satisfied that in reasonable probability the evidence has not been altered in any material respect, he may permit its introduction.” (citation omitted)). There is a strong public policy for making all information filed in a quasi-judicial administrative proceeding available to the public, especially in an *inter partes* review, which determines the patentability of a claim in an

issued patent. It is within the Board’s discretion to assign the appropriate weight to be accorded to evidence.

Although Messrs. Sussell and Sadofsky, prior to the critical date, personally did not post or review the particular version of Kantor relied upon by EMC (Ex. 1004), they nevertheless have sufficient personal knowledge and working experience to provide competent testimony to establish the publication and authentication of Kantor. *See Hall*, 781 F.2d at 899; *Wyer*, 655 F.2d at 226; *Bayer*, 568 F.2d at 1361. Notably, Mr. Sussell, the co-founder and system operator of the Invention Factory Bulletin Board System, testifies that Dr. Kantor released the first version of his software on the Invention Factory Bulletin Board System in the 1980s, and the system continuously utilized and hosted current versions of the software and user manuals. Ex. 1049 ¶¶ 3, 13, 15. Mr. Sussell also testifies that the Invention Factory Bulletin Board System advertised Dr. Kantor’s software to its users by including information about Dr. Kantor’s software on the “Welcome” screen, and made FWKCS Zip file—a zip file that contains both the software and user manual, Kantor—publicly accessible and available under four different directories. *Id.* ¶ 18. According to Mr. Sussell, the Invention Factory Bulletin Board System had over 3,000 subscribers, in the 1993 timeframe, and all of the users had the capability to perform keyword searches to retrieve FWKCS Zip file. *Id.* ¶¶ 6, 21.

Although we are cognizant that electronic documents generally are not self-authenticating, it has been recognized that “[t]o authenticate printouts from a website, the party proffering the evidence must produce

some statement or affidavit from someone with knowledge of the website . . . for example a web master or someone else with personal knowledge would be sufficient.” *St. Luke’s Cataract and Laser Institute v. Sanderson*, 2006 WL 1320242, *2 (M.D. Fla. 2006) (quoting *In re Homestore.com, Inc. Sec. Litig.*, 347 F. Supp. 2d 769, 782 (C.D. Cal. 2004)) (internal quotation marks omitted); Ex. 2024; *see also Market-Alerts Pty. Ltd. v. Bloomberg Finance L.P.*, 922 F. Supp. 2d 486, 493, n.12 (D. Del. 2013) (citing *Keystone Retaining Wall Sys., Inc. v. Basalite Concrete Prods., LLC*, 2011 WL 6436210, *9 n.9 (D.Minn. 2011)) (Documents generated by a website called the Wayback Machine have been accepted generally as evidence of prior art in the patent context.); *U.S. v. Bansal*, 663 F.3d 634, 667-68 (3d. Cir. 2011) (concluding that the screenshot images from the Internet Archive were authenticated sufficiently under Federal Rule of Evidence 901(b)(1) by a witness with personal knowledge of its contents, verifying that the screenshot the party seeks to admit are true and accurate copies of Internet Archive’s records).

Here, Mr. Sadofsky, who is a technology archivist and software historian and currently is an archivist for the Internet Archive, testifies that he launched the website textfiles.com and a subdomain cd.textfiles.com to collect software, data files, and related materials from Bulletin Board Systems. Ex. 1077 ¶¶ 9-11. According to Mr. Sadofsky, textfiles.com and cd.textfiles.com are dedicated to preserving, archiving, and providing free access to unaltered historical software programs and information that initially were made available on the Bulletin Board Systems. *Id.*

Mr. Sadofsky states that he previously archived the FWKCS Zip file (FWKCS122.ZIP) that contains Dr. Kantor’s software and user manual to cd.textfiles.com from his own copy of the *Simtel MSDOS Archive*, October 1993 Edition, Walnut Creek CD-ROM. *Id.* ¶ 14 (citing Ex. 1048).

Mr. Sadofsky also testifies that he personally verified the authenticity of Kantor—version 1.22, the version relied upon by EMC (Ex. 1004)—by comparing it with the “1993 archived” version and determined that Kantor is identical to the “1993 archived” version. Ex. 1077 ¶¶ 13-15. Mr. Sadofsky confirms that the source file of the “1993 archived” version has a timestamp of August 10, 1993, at 1:22 AM. *Id.* ¶ 16; *see also* Ex. 1087 ¶¶ 10-11; Ex. 2014 ¶ 5. Mr. Sadofsky concludes that Kantor was publicly accessible prior to the critical date. Ex. 1077 ¶¶ 13, 16. Therefore, we agree with EMC that Kantor has been authenticated sufficiently to warrant its admissibility under Federal Rules of Evidence 901(b)(1), (b)(3), and (b)(4).

In addition, we agree with EMC that Kantor also has been authenticated as an “ancient document” under Federal Rule of Evidence 901(b)(8).⁵ Opp. 7. Kantor is “at least 20 years old and can be found in . . . an October 1993 *Simtel* CD-ROM – a place where an authentic 20-year old document distributed through a [Bulletin Board System] would likely be.”

⁵ Fed. R. Evid. 901(b)(8). Evidence About Ancient Documents or Data Compilations. For a document or data compilation, evidence that it:
(A) is in a condition that creates no suspicion about its authenticity;
(B) was in a place where, if authentic, it would likely be; and
(C) is at least 20 years old when offered.

Id.; Ex. 1077 ¶¶ 7-8; *see also* Fed. R. Evid. 901(b)(8) 2012 Adv. Comm. Note (“The familiar ancient document rule of the common law is extended to include data stored electronically or by other similar means.”). Moreover, testimony of Messrs. Sussell and Sadofsky has established sufficiently that Kantor is in a condition that creates no suspicion about its authenticity. Exs. 1049, 1077, 1087.

PersonalWeb does not present sufficient or credible evidence to the contrary. Based on the evidence before us, we determine that Kantor has been authenticated sufficiently under Federal Rules of Evidence 901(b)(1), (b)(3), (b)(4), and (b)(8) to warrant its admissibility.

PersonalWeb’s hearsay argument regarding Kantor also is unavailing. As EMC notes (Opp. 8), a “prior art document submitted as a ‘printed publication’ under 35 U.S.C. § 102(a) is offered simply as evidence of what it described, not for proving the truth of the matters addressed in the document.” *See, e.g., Joy Techs., Inc. v. Manbeck*, 751 F. Supp. 225, 233 n.2 (D.D.C. 1990), *judgment aff’d*, 959 F.2d 226 (Fed. Cir. 1992); Fed. R. Evid. 801(c) 1997 Adv. Comm. Note (“If the significance of an offered statement lies solely in the fact that it was made, no issue is raised as to the truth of anything asserted, and the statement is not hearsay.”). Therefore, Kantor is not hearsay under Federal Rule of Evidence 801(c).

We further agree with EMC that the posted date of “1993 August 10” or the copyright date of “1988-1993” on the Title page of Kantor is not a basis for excluding Kantor, as testimony from Messrs. Sussell and Sadofsky sufficiently establishes that Kantor existed as of August 10, 1993, prior to

the critical date. Opp. 8. More importantly, the computer-generated timestamp—August 10, 1993, at 1:22 AM—of the “1993 archived” version of Kantor (Ex. 1077 ¶¶ 14-15; Ex. 1087 ¶¶ 10-11; Ex. 2014 ¶ 5) also independently corroborates Kantor’s existence as of August 10, 1993. *See, e.g., U.S. v. Khorozian*, 333 F.3d 498, 506 (Fed. Cir. 2003) (concluding that an automatically generated time stamp on a fax was not a hearsay statement because it was not uttered by a person). Accordingly we are not persuaded that PersonalWeb has presented a sufficient basis to exclude Kantor, as impermissible hearsay.

For the foregoing reasons, we decline to exclude Kantor.

2. Documents Corroborating Witnesses’ Knowledge and Recollections

PersonalWeb asserts that certain documents submitted by EMC (Exs. 1046-1048, 1051-1054, 1073, 1074, 1079-1081) and the declarations of Messrs. Sussell and Sadofsky (Exs. 1049, 1077, 1087) regarding these documents should be excluded because the documents have not been authenticated properly and are inadmissible hearsay. PO Mot. 6-9.

PersonalWeb argues that EMC “has not established that any of these documents existed prior to the critical date, and no witness has personal knowledge of their alleged existence prior to April 11, 1995.” *Id.* at 7.

PersonalWeb further maintains that the documents that are Exhibits 1052, 1053, 1073, and 1074 are irrelevant, prejudicial, and confusing, as they discuss a version of Kantor different than the version relied upon by EMC (version 1.22, Ex. 1004). *Id.* at 8-9.

EMC responds that its witnesses provided those documents to corroborate their independent knowledge and recollections. Opp. 10. EMC asserts that the documents have been authenticated under Federal Rules of Evidence 901-902 and fall within a hearsay exception under Federal Rules of Evidence 803-807. *Id.* at 10-12. We are persuaded by EMC's arguments.

As the movant, PersonalWeb has the burden of proof to establish that it is entitled to the requested relief. 37 C.F.R. § 42.20(c). As discussed previously, we disagree with PersonalWeb that documents cannot be authenticated without direct testimony from the author or a witness who actually reviewed the documents prior to the critical date. *See* Fed. R. Evid. 901(a). Significantly, PersonalWeb's motion does not contain sufficient explanations why each document should be excluded. For instance, PersonalWeb does not explain adequately why the declaration of Mr. Sussell (Ex. 1049 ¶¶ 6, 8, 18, 27) is not sufficient to authenticate Exhibits 1051-1054, 1073, and 1074, or why the declarations of Mr. Sadofsky (Ex. 1077 ¶¶ 7-17; Ex. 1087 ¶¶ 10-16) are not sufficient to authenticate Exhibits 1046-48 and 1079-1081. *See* Fed. R. Evid. 901(b)(1).⁶ Nor does PersonalWeb explain sufficiently why the following documents are not self-authenticated: (1) Exhibits 1046-1048 and 1051 – articles that have LexisNexis® trade inscriptions; (2) Exhibits 1073 and 1074 – Usenet newsgroup periodicals that have Usenet trade inscriptions; and (3) Exhibit

⁶ Fed. R. Evid. 901(b)(1). Testimony of a Witness with Knowledge.
Testimony that an item is what it is claimed to be.

1048 – a photograph of the *Simtel MSDOS Archive*, October 1993 Edition, Walnut Creek CD-ROM, that has Simtel trade inscriptions. *See* Fed. R. Evid. 902(6)-(7).⁷

In its motion, PersonalWeb fails to identify, specifically, the textual portions of the aforementioned exhibits that allegedly are being offered for the truth of the matter asserted, yet seeks to exclude the entirety of each exhibit. The burden should not be placed on the Board to sort through the entirety of each exhibit and determine which portion of the exhibit PersonalWeb believes to be hearsay. Rather, PersonalWeb should have identified, in its motion, the specific portions of the evidence and provided sufficient explanations as to why they constitute hearsay. Furthermore, PersonalWeb does not explain adequately why the declarations of Messrs. Sussell and Sadofsky do not provide the proper foundation and corroboration for the exhibits.

To the extent PersonalWeb relies upon the same arguments with respect to Kantor for excluding the aforementioned exhibits, we have addressed those arguments above and determined that they are unavailing. We also

⁷ Fed. R. Evid. 902. Evidence that Is Self-Authenticating
The following items of evidence are self-authenticating; they require no extrinsic evidence of authenticity in order to be admitted:

. . . .

(6) Newspapers and Periodicals. Printed material purporting to be a newspaper or periodical.

(7) Trade Inscriptions and the Like. An inscription, sign, tag, or label purporting to have been affixed in the course of business and indicating origin, ownership, or control.

agree with EMC that the exhibits concerning prior versions of Kantor are relevant, and not prejudicial or confusing as alleged by PersonalWeb, because such circumstantial evidence provides context and corroboration for the witnesses' independent knowledge and recollection.

Furthermore, we are not persuaded that the declarations of Messrs. Sussell and Sadofsky (Exs. 1049, 1077, 1087) should be excluded. As we discuss above and below in the next section, they have sufficient personal knowledge and working experience to provide competent testimony to establish the publication and authentication of Kantor. The documents they cite serve to corroborate their independent knowledge and recollection.

For the foregoing reasons, PersonalWeb has not presented a sufficient basis to exclude Exhibits 1046-1048, 1051-1054, 1073, 1074, and 1079-1081, as well as the declarations of Messrs. Sussell and Sadofsky (Exs. 1049, 1077, 1087) concerning those exhibits.

3. Declarations of Messrs. Sussell and Sadofsky

PersonalWeb argues that the declarations of Messrs. Sussell and Sadofsky (Exs. 1049, 1077, 1087) should be excluded as hearsay under Federal Rule of Evidence 801 and inadmissible under Federal Rules of Evidence 802-807 for lack of foundation and personal knowledge, and Federal Rule of Evidence 702 as improper testimony, because the witnesses personally did not review Kantor (Ex. 1004) and Simtel (Ex. 1048) prior to the critical date. PO Mot. 9. PersonalWeb also argues that Messrs. Sussell and Sadofsky “are not qualified experts in the field.” *Id.* at 11.

PersonalWeb further alleges that Mr. Sadofsky's deposition (Ex. 2013, 30,

66) should be excluded, as it was responsive to a leading question and non-responsive to the question. *Id.*

EMC responds that the testimony of Messrs. Sussell and Sadofsky should not be excluded, because their testimony is based on their own personal knowledge and recollection, and the documents they cite serve to corroborate their independent knowledge and recollection. Opp. 13. EMC further explains that the witnesses have described thoroughly the underlying facts, and, therefore, the testimony should be admitted as relevant under Federal Rules of Evidence 401-402, supported by personal knowledge and foundation under Federal Rule of Evidence 602, and proper opinion testimony under Federal Rules of Evidence 701-703. *Id.* We find EMC's contentions have merit.

PersonalWeb's arguments rest on the erroneous premise that EMC's witnesses must have reviewed Kantor or Simtel personally prior to the critical date in order to provide competent testimony regarding Kantor or Simtel. As discussed previously, it is well settled that it is not necessary for the witnesses to have reviewed the reference personally prior to the critical date in order to establish publication. *See, e.g., Wyer*, 655 F.2d at 226.

Although Messrs. Sussell and Sadofsky are not experts related to the claimed subject matter of the '544 patent, each witness nevertheless has sufficient personal knowledge and working experience to provide competent testimony. *See Hall*, 781 F.2d at 899. Mr. Sussell was the co-owner and system operator of the Invention Factory Bulletin Board System from 1983 to 1996. Ex. 1049 ¶ 3. Mr. Sussell's testimony is based on his personal

knowledge of the relevant facts related to the Invention Factory Bulletin Board System and Kantor. *Id.* at ¶ 2. Notably, Dr. Kantor specifically thanked Mr. Sussell in his user manual for hosting Dr. Kantor’s software FWKCS and for Mr. Sussell’s role in its development. Ex. 1004, 3 (“To Michael Sussell, sysop of The Invention Factory (R), home board for the support of FWKCS, for bringing the problem of duplicate files to my attention and for his help in testing”); *id.* at 6 (“When Michael Sussell, sysop of The Invention Factory (R) in New York, brought to my attention the problem of duplicate files with different names, these concepts provided valuable insight into how one might proceed.”).

Mr. Sadofsky is a technology archivist and software historian, and, currently, works “for the Internet Archive, a non-profit digital library offering free universal access to books, movies, and music, as well as 342 billion archived webpages available through the Wayback Machine service.” Ex. 1077 ¶ 3. Mr. Sadofsky also “directed the film, *The BBS Documentary*, an eight-episode documentary about the subculture born from the creation of the [Bulletin Board System].” *Id.* at ¶ 4. Mr. Sadofsky’s testimony is based on his personal knowledge of the relevant facts related to Kantor and the “1993 archived” version of Kantor. *Id.* at ¶ 2; Ex. 1087 ¶ 2. For example, Mr. Sadofsky personally verified the authenticity of Kantor by comparing it with the “1993 archived” version, and determined that Kantor—version 1.22, the version relied upon by EMC (Ex. 1004)—is identical to the “1993 archived” version. Ex. 1077 ¶¶ 14-15.

Upon review of the evidence on the record, we agree with EMC that both Messrs. Sussell and Sadofsky have disclosed sufficient underlying facts to support their testimony. For instance, the computer-generated timestamp—August 10, 1993, 1:22 AM—associated with the “1993 archived” version of Kantor corroborates their testimony regarding Kantor’s existence as of August 10, 1993. Ex. 1077 ¶¶ 14-15; Ex.1087 ¶¶ 10-11; Ex. 2014 ¶ 5.

As to Mr. Sadofsky’s deposition (Ex. 2013, 30, 66), PersonalWeb does not explain sufficiently why that testimony should be excluded. PO Mot. 11. Moreover, Mr. Sadofsky’s deposition (Ex. 2013, 30, 66) is consistent with his direct testimony (Ex. 1077 ¶¶ 14-16), and, therefore, it would not prejudice PersonalWeb even if such evidence is not excluded.

For the foregoing reasons, PersonalWeb has not presented a sufficient basis to exclude the declarations of Messrs. Sussell and Sadofsky (Exs. 1049, 1077, 1087) and Mr. Sadofsky’s deposition (Ex. 2013, 30, 66).

4. Clark’s Rebuttal Declaration

PersonalWeb alleges that Dr. Clark’s rebuttal declaration (Ex. 1088 ¶¶ 26-27, 30) should be excluded, because it is irrelevant, prejudicial, confusing, and beyond the scope of this proceeding. PO Mot. 11-14. In particular, PersonalWeb argues that what Kantor’s software allegedly could do is irrelevant because an *inter partes* review is limited to printed publications and patents. *Id.* at 11-12. PersonalWeb also contends that Dr. Clark’s rebuttal declaration (Ex. 1088 ¶¶ 26-27) contradicts his prior deposition (Ex. 2015, 55, 59, 66-67) and constitutes new evidence that

should have been presented earlier. *Id.* at 12-14.

EMC counters that Dr. Clark’s rebuttal declaration (Ex. 1088 ¶¶ 26-27, 30) regarding what Kantor’s software “could” do is relevant and admissible. Opp. 13-14. According to EMC, Dr. Clark’s rebuttal declaration (Ex. 1088 ¶ 30) was submitted in response to PersonalWeb’s argument that “Kantor’s software reads part identifiers (CRC values) from a zip file instead of generating them” (PO Resp. 20-22; Ex. 2016 ¶¶ 49-50). *Id.* EMC points out that Dr. Clark’s rebuttal declaration merely explains “what a person of skill in the art, reading the Kantor reference, would [have understood] the reference to disclose about Kantor’s software, including the software’s disclosed capabilities to generate CRC values when it cannot look them up.” *Id.* at 14 (citing Ex. 1088 ¶ 30).

Having reviewed PersonalWeb’s patent owner response and Dr. Clark’s rebuttal declaration, we determine that Dr. Clark’s testimony is reasonable rebuttal evidence in light of PersonalWeb’s arguments submitted in its patent owner response. Notably, Dr. Clark’s rebuttal declaration responds appropriately to the issue raised by PersonalWeb (PO Resp. 20-22; Ex. 2016 ¶¶ 49-50)—whether Kantor, a software user manual, describes generating a hash of hashes, as required by claim 1, when Kantor describes how the software calculates a zip-file contents signature. PersonalWeb has not demonstrated sufficiently that Dr. Clark’s rebuttal testimony is irrelevant or exceeds the proper scope of reply evidence.

We also are not persuaded by PersonalWeb’s argument that Dr. Clark’s rebuttal declaration (Ex. 1088 ¶¶ 26-27) contradicts his earlier

testimony (Ex. 2015, 55, 59, 66-67). Rather, we agree with EMC (Opp. 14-15) that Dr. Clark’s rebuttal testimony that “zip files are not *always* compressed” (Ex. 1088 ¶¶ 26-27) is consistent with his earlier testimony that the inner files of a zip file are compressed *typically* (Ex. 2015, 55, 59, 66-67). Moreover, Dr. Clark’s testimony (Ex. 1088 ¶¶ 26-27) is reasonable rebuttal evidence in light of the evidence submitted by PersonalWeb in support of its patent owner response. Dr. Clark merely points out in his rebuttal declaration that PersonalWeb’s evidence also shows that zip files are not *always* compressed. Ex. 1088 ¶ 26 (citing Ex. 2007, 3 (The zip file format defines seven compression methods, including “Compression method 0,” which does not compress the file.); Ex. 1084, 262 (Dr. Dewar agrees that “the zip file standard allows for uncompressed files.”)).

For the foregoing reasons, we decline to exclude Dr. Clark’s rebuttal declaration (Ex. 1088).

III. CONCLUSION

EMC has met its burden of proof by a preponderance of the evidence in showing that claim 1 the ’544 patent is unpatentable based on the following grounds of unpatentability:

| Claim | Basis | References |
|-------|----------|---------------------|
| 1 | § 102(e) | Woodhill |
| 1 | § 102(b) | Kantor |
| 1 | § 103(a) | Kantor and Woodhill |

IV. ORDER

In consideration of the foregoing, it is
ORDERED that claim 1 of the '544 patent is held unpatentable;
FURTHER ORDERED that EMC's Motion to Exclude Evidence is
dismissed;

FURTHER ORDERED that PersonalWeb's Motion to Exclude
Evidence is *denied*; and

FURTHER ORDERED that because this is a final written decision,
parties to the proceeding seeking judicial review of the decision must
comply with the notice and service requirements of 37 C.F.R. § 90.2.

Case IPR2013-00084
Patent 7,945,544 B2

PETITIONER:

Peter M. Dichiara, Esq.
David L. Cavanaugh, Esq.
WILMER CUTLER PICKERING HALE & DORR LLP
peter.dichiara@wilmerhale.com
david.cavanaugh@wilmerhale.com

PATENT OWNER:

Joseph A. Rhoa, Esq.
Updeep S. Gill, Esq.
NIXON & VANDERHYE P.C.
jar@nixonvan.com
usg@nixonvan.com