Digital Humanities in Biblical Studies and Theology

Sarah Yardney*, Sandra R. Schloen, Miller Prosser

New Digital Tools for a New Critical Edition of the Hebrew Bible

<https://doi.org/10.1515/opth-2019-0006>

Received January 14, 2019; accepted April 11, 2019

Abstract: This article describes the digital edition of the Hebrew Bible: A Critical Edition (HBCE), which is being produced as part of a project called Critical Editions for Digital Analysis and Research (CEDAR) at the University of Chicago. We first discuss the goals of the HBCE and its requirements for a digital edition. We then turn to the CEDAR project and the advances it offers, both theoretical and technological. Finally, we present an illustration of how a reader might use the digital HBCE to interact with the biblical text in innovative ways.

Keywords: Textual criticism, digital text, textual encoding, database

1 Introduction

Critical Editions for Digital Analysis and Research (CEDAR) is a new platform for textual research, currently under development at the University of Chicago, that will host the digital version of The Hebrew Bible: A Critical Edition (HBCE).¹ Despite the great expansion of our textual data since the earliest works of biblical text-criticism, there have been almost no conceptual advances in the tools available for studying that data.² Word processors have made publishing easier, and digital imaging techniques can produce clearer and more detailed photographs, but for the most part scholars are still working with flat, static text and flat, static images.³ The tools offered by CEDAR, made possible by a new approach to encoding texts, thus represent a significant development in the history of textual criticism.

In this article, we first discuss the goals of the HBCE and its requirements for a digital edition. We then turn to the CEDAR project and the advances, both theoretical and technological, that make it ideally suited for the digital HBCE. Finally, we present an illustration of how a reader might use the digital HBCE to interact with the biblical text in innovative ways.

¹ Ronald Hendel, general editor (Society of Biblical Literature, 2015–). Each book of the Hebrew Bible has its own editor and will be published individually. The first completed volume is Michael V. Fox’s edition of Proverbs (SBL, 2015). A full description of the project’s goals and methodology can be found in Hendel, *Steps to a New Edition of the Hebrew Bible*, 15–63. See also the HBCE website at https://www.sbl-site.org/HBCE/HBCE_About.html [accessed 10 December 2018].

² The earliest known work of biblical textual criticism is Origen’s third-century *Hexapla*, which compared the Hebrew with several Greek translations in multiple columns; see Tov, *Textual Criticism of the Hebrew Bible*, 145–46. The shortcomings of currently available tools are discussed in Yardney, Schloen, and Prosser, “A Preliminary Report from the CEDAR Project.”

³ There are a few exceptions, notably the Scripta Qumranica Electronica (SQE), which is developing digital tools for reconstructing Dead Sea Scrolls; see Brown-deVost, “Scripta Qumranica Electronica (2016–2021).” In the realm of images, a technique known as Reflectance Transformation Imaging (RTI) produces dynamic images of objects, including texts, that can be viewed with different light sources that illuminate physical features of the object. Unfortunately, the SQE is currently available only to those working on the project, and RTI images are still relatively few in number and not easily accessed.

***Corresponding author: Sarah Yardney**, University of Chicago, Chicago, United States of America;

E-mail: syardney@uchicago.edu

Sandra R. Schloen, Miller Prosser, University of Chicago, Chicago, United States of America

2 The goals and requirements of the HBCE

The HBCE is the first eclectic edition of the Hebrew Bible to be produced.⁴ The project aims “to approximate the corrected archetype of each biblical book.”⁵ By “archetype,” the editors mean the latest common ancestor of all the variant manuscripts. This ancestor can be inferred, with due caution and epistemological humility, from the textual witnesses of those manuscripts. The archetype – the *latest* common ancestor – differs from the original in what we can know about it: the archetype can be inferred from the available data; the original cannot. Further, the archetype may represent a different textual state from the original, although we cannot know with certainty because we have no access to the original. The archetype, and not the original, is the target of the HBCE. The HBCE intends to approximate the *corrected* archetype because the archetype “will contain scribal errors that can sometimes be detected and corrected. For instance, distinctive scribal errors that are in *all* the manuscripts derive from the archetype. If we can discern these errors, we are obliged as textual critics to correct them.”⁶ Finally, the editors aim to *approximate* the corrected archetype because they recognize the potential fallibility of their judgments as well as the methodological impossibility of verifying the reconstruction of a hypothetical text.

Each volume of HBCE will be offered in both print and digital versions. The digital version will be an interactive polyglot, comprising the HBCE critical text and commentary as well as the evidence used to construct the critical edition.⁷ The digital version will thus allow scholars to engage with the textual data in ways not possible in print. Instead of being limited by a critical apparatus showing the variants deemed salient by an editor, scholars will choose which manuscripts they want to compare and which kinds of variants they want to see. They will also have access to full transcriptions and images of those manuscripts so that variants can be understood in context. In addition, scholars will be able to perform their own text-critical comparisons of manuscripts, both within and across languages.

The digital HBCE will be produced as part of a new project called Critical Editions for Digital Analysis and Research, or CEDAR. Based in the Oriental Institute at the University of Chicago, CEDAR is a collaboration between computer scientists and textual scholars. The project has created new ways of encoding manuscripts that enable deeper and more comprehensive study of texts, their histories, and their relationships.⁸ Although this article focuses on CEDAR’s application to biblical studies, the tools we are developing are generic enough to be used for any textual corpus.⁹

CEDAR is built within a high-performance XML graph database called the Online Cultural and Historical Research Environment, or OCHRE.¹⁰ OCHRE was originally developed to record and manage archaeological data, but it now houses over fifty projects across a wide range of disciplines in the humanities, social sciences, and hard sciences. For over a decade, OCHRE has been tested by philologists recording texts

⁴ While eclectic editions – editions which represent the best version of a text as determined by an editor – are the norm in New Testament studies, as well as in other disciplines in the humanities, scholarly editions of the Hebrew Bible have always reproduced a single manuscript and provided variant readings in a critical apparatus. The standard scholarly edition of the New Testament is Nestle-Aland’s 28th edition of the *Novum Testamentum Graece* (Deutsche Bibelgesellschaft, 2012), an eclectic edition. The widely available *The Greek New Testament: SBL Edition* (SBL, 2010) is also eclectic. The standard edition of the Hebrew Bible, by contrast, *Biblica Hebraica Stuttgartensia* (Deutsche Bibelgesellschaft, 1977), as well as the new *Biblia Hebraica Quinta* (Deutsche Bibelgesellschaft, 2004–), which will replace it, are both diplomatic editions which reproduce the Leningrad Codex. The other major scholarly edition of the Hebrew Bible, the *Hebrew University Bible* (1956–), instead reproduces the Aleppo Codex.

⁵ Hendel, *New Edition*, 21.

⁶ *Ibid.*, 23; italics original.

⁷ “Method & Theory,” https://www.sbl-site.org/HBCE/HBCE_Method.html [accessed 10 December 2018].

⁸ A full description of CEDAR can be found at <https://cedar.uchicago.edu>.

⁹ In order to ensure CEDAR’s flexibility, we are working simultaneously on three test cases in addition to the Hebrew Bible: the Sumerian *Epic of Gilgamesh* (ca. 2100 BCE), Shakespeare’s *The Taming of the Shrew* (1623), and two examples of middle Bengali poetry, Ālāol’s (fl. 1651–1671 CE) versified Sufi romance *Sayphulmuluk and Badiujjāmāl*, and an anthology of 308 lyric poems compiled by Viśvanāth Cakravartī “Harivallabh” entitled *Kṣaṇadāgitacintāmaṇi* (ca. 1704 CE).

¹⁰ Descriptions of OCHRE can be found in Schloen and Schloen, *UCHRE: An Online Cultural and Historical Research Environment* and Schloen and Schloen, “Beyond Gutenberg.” See also “UCHRE,” <https://ochre.uchicago.edu/page/ochre> [accessed 20 December 2018].

with highly complex writing systems. Together, some 25,000 texts have been entered in the database, in languages as varied as Elamite, Assyrian, Egyptian, Greek, Ugaritic, Aramaic, and Demotic. CEDAR is the first text-critical project in OCHRE, but the database has several features that make it well-suited for this kind of work. In particular, OCHRE enables multiple overlapping hierarchies by using an item-based ontology in which representations of hierarchical structures are stored as separate items from the data in those hierarchies. The OCHRE data model is also highly granular, allowing scholars to atomize their data into the most minimal meaningful parts for analysis and manipulation. Finally, implementing CEDAR in OCHRE allows scholars to have a great deal of flexibility both in what data they choose to capture and in how they describe the relationships among their data. These features allow for significant conceptual and computational advances in representing and interacting with digital texts.¹¹

3 Conceptual advances in CEDAR

CEDAR implements a new approach to encoding text that might be called the database model.¹² In the CEDAR project, although we are creating tools to represent and study texts, there are no “texts” per se stored in the database. That is to say, the database does not store linear strings of characters comprising whole compositions. Instead, each verse, each line, each word, each individual character, and even each individual diacritical mark is stored as its own discrete XML document. The various hierarchies in which they can be arranged are stored individually as separate XML documents. When a user asks to view a text, OCHRE accesses the appropriate hierarchies, gathers the database items associated with those hierarchies, assembles these items into words and whatever other structure is defined by the hierarchies (e.g., lines/columns/pages, verses/chapters/books), and displays an organized, formatted view of the text in the linear strings of characters that readers are accustomed to seeing. There is a distinction, in other words, between how the user reads the data and how the data is organized in the database.

This distinction is not maintained in the dominant approach to encoding texts in the humanities, which might be called the document model.¹³ The document model both displays and organizes data in the same pre-digital structure of physical documents, in which characters are assigned a fixed position in one or two dimensions. As Schloen and Schloen note,

some kinds of scholarly work can be done using this method, but it imposes unnecessary limits if it is used as the primary means of representing a text. It deprives scholars of the power to express in digital form many of the conceptual distinctions they employ routinely in the course of their work. They end up imitating the position-dependent structure of a pre-digital medium rather than exploiting the potential of the digital medium to represent their information in a more effective way. As a result, they fail to capture in an explicit, searchable form the different ways a given text has been read and annotated, making it difficult to use computational methods to compare and analyze the various interpretations. And it is precisely those texts which are open to many different readings that are of greatest interest to scholars in the humanities, who spend considerable time tracing the history of textual interpretation and the interconnections within and among texts.¹⁴

Digitization of texts should offer scholars new and more effective ways of doing their work, but the document model fails to take advantage of that potential because it reproduces the design of print media. The database model, on the other hand, can capture multiple interpretations and versions of a given text, which are precisely what scholars in the humanities are often most interested in studying.

¹¹ Another widely used approach is to encode texts in XML documents marked up according to the standards of the Text Encoding Initiative (TEI); see <http://www.tei-c.org/> [accessed 20 December 2018]; Cummings, “The Text Encoding Initiative and the Study of Literature.” XML is sometimes mistaken to be synonymous with TEI, but they are in fact distinct. XML is a markup language, while TEI is a set of guidelines for standardizing markup tags within that language. OCHRE uses XML but not TEI. The research questions we want to ask in the CEDAR project require a highly atomized, normalized data model which TEI is not intended to provide. TEI remains a useful way to capture many aspects of transcribed texts, however, and data can be exported from OCHRE in TEI format.

¹² The database model and the document model (see below) were first described in Schloen and Schloen, “Beyond Gutenberg.”

¹³ *Ibid.*

¹⁴ *Ibid.*, 4.

The database model is especially powerful for textual criticism because the same database items can be reused in different versions of the same text. In OCHRE, each attested character and diacritical mark is encoded as a separate database item. For example, in *bərē'sīt*, the first word of Genesis in Hebrew, the *bêt* is an XML document, the *dāgēš* is an XML document, the *šəwā'* is an XML document, and so on (see Figure 1). Combinations of these items are then assigned to hierarchies that represent individual manuscripts. When the user asks to view the various manuscripts of Genesis, OCHRE generates them on-the-fly by gathering and arranging the same underlying database items according to each manuscript's hierarchies. The user sees separate texts, but within the database, each “text” is simply a recombination of the same building blocks.

Figure 1: The reuse of database items in multiple contexts

This structure means that instead of existing in one or two dimensions, as in the document model, texts in the database model exist in three dimensions. Rather than being constrained to setting them side by side for comparison, we can stack one on top of the other.

Figure 2: Two-dimensional text in the document model vs. three-dimensional text in the database model

Texts become like transparency sheets, and we can compare texts by layering one on top of the other to see where they line up. To add another text to the database, we simply assign the appropriate already-existing characters to another set of hierarchies; the new text is instantly related to other texts already in the database. By contrast, a comparison of multiple manuscripts in the document model would require the creation of many duplicate character strings with links between them. Instead of one single *bêt* reused in every instance of *bərē'sīt* in every manuscript, as we have in the database model, we would have as many *bêt*'s as we have manuscripts, with a giant web of links to relate them.

Figure 3: Comparing multiple texts in the document model and in the database model

4 Computational strategies in CEDAR

We employ a number of new computational strategies that result in the conceptual advances discussed above. First, CEDAR makes use of OCHRE's ability to encode overlapping hierarchies. Textual data are organized into two overlapping hierarchies: an epigraphic hierarchy, which is arranged by material and graphical features such as pages, columns, lines, and characters; and a discourse hierarchy, which is arranged by discursive features such as sentences, phrases, and words (or, for a biblical text, books, chapters, verses, and words).¹⁵ Each word in the discourse hierarchy is comprised of characters from the related epigraphic hierarchy.

epigraphic hierarchy:
material and graphical features

discourse hierarchy:
discursive features

Figure 4: The epigraphic and discourse hierarchies in OCHRE

¹⁵ This structure accommodates the important observation made by Dino Buzzetti (Buzzetti, “Digital Representation and the Text Model,” 76) that the expression of the text and the content of the text be encoded separately.

Second, texts in CEDAR are highly atomized, as evident in Figure 4. Each character and diacritical mark is a separate database item – more specifically, a separate XML document – with its own universally unique identification number (UUID). For example, in Figure 4, the first character in the epigraphic hierarchy is a *bêt* (the underlying character) with a *dāgēš* and a *šēwā*’ (the diacritical dots in the middle of and below the character). In CEDAR, that epigraphic unit is broken down into its three constituent elements, each of which is itself a separate database item. Each database item can be used and marked with scholarly interpretation independent of the others. While this degree of atomization may seem extreme, it is required for the work of text-critical scholars, who attend to differences at this level of detail.

Third, our strategy of atomization makes it possible to compare manuscripts in great detail and with great flexibility. In the example in Figure 5, three different options are attested for the diacritics on the *iota* in the word *εἰδεν* in Gen 1:4. We represent these options by creating three different epigraphic units, but nested within each is the same *iota*. We are not forced to normalize spelling, but the different options are still recognized as the same *iota* because of the reuse of the underlying character. Users can then choose whether they want CEDAR to highlight differences in diacritics or not.

= same database item

Figure 5: Reuse of database items in multiple combining characters

Atomization also allows comparison between manuscripts with diacritics and manuscripts without, while still avoiding duplication of character strings. We can use the *iota* in Figure 5 to encode a Greek manuscript that attests no diacritics at all by adding the *iota* alone to that manuscript’s hierarchy, leaving the diacritics behind. The reuse of the database item creates alignment between that unaccented *iota* and any of the accented options.

Fourth, whenever possible, we create relationships between texts by reusing items in the database rather than by creating new items and linking them.¹⁶ This strategy has a number of benefits: it avoids data redundancy and eliminates the proliferation of links that would otherwise be needed to align content, it reduces the copying errors that inevitably occur whenever another version of a text is produced by human hands, and adding the 100th version of a text and relating it to all the others is no more difficult than adding the second version.

The principle of reuse is achieved through what we call a content pool. A content pool is the set of all readings attested in physical manuscripts and critical editions of a given text. It represents not any one

¹⁶ The one circumstance in which we do not reuse database items is when we want to assign properties to a character that obtain only in one manuscript – for example, when we want to mark that a character is damaged, written above the line, ligatured, partially erased, etc. In these cases, we create a copy of the database item and mark the copy. If we were to assign a property to a reused character, that property would propagate in all manuscripts in which that character is used. Copied characters have no impact on CEDAR’s ability to perform comparisons between manuscripts.

particular text but rather all of the possibilities of a textual tradition organized into hierarchies. It is thus an abstraction. For example, in Gen 1:6, two Greek manuscripts have omega as the final character in γεννητω, while another has omicron (γεννητο).¹⁷ The Greek content pool in this verse thus reads γεννητωο. Note that the content pool is not simply the alphabet for a given language, and it is not the case that every single *bêt* in the Hebrew text of Genesis is the same *bêt*: rather, every instance of the initial *bêt* in the first word of Genesis is the same. Each subsequent *bêt* is a different database item. After the content pool has been created, characters are then selected from it to build individual manuscripts, which we call local texts.¹⁸ Local texts are essentially subsets of data selected from the content pool.

The Hebrew content pool for Gen 1:9

וַיֹּאמֶר אֱלֹהִים יִקְוֹא וְהָיָה מִיֶּסֶד לַשָּׁמַיִם וְהָיָה יְהִי יְהִי
וְיִקְוֹה מִיֶּסֶד הַשָּׁמַיִם לַמְּקוֹה וְהָיָה יְהִי יְהִי

The Leningrad Codex

וַיֹּאמֶר אֱלֹהִים יִקְוֹא וְהָיָה מִיֶּסֶד לַשָּׁמַיִם וְהָיָה יְהִי יְהִי
וְיִקְוֹה מִיֶּסֶד הַשָּׁמַיִם לַמְּקוֹה וְהָיָה יְהִי יְהִי

4Q2

וַיֹּאמֶר אֱלֹהִים יִקְוֹא וְהָיָה מִיֶּסֶד לַשָּׁמַיִם וְהָיָה יְהִי יְהִי
וְיִקְוֹה מִיֶּסֶד הַשָּׁמַיִם לַמְּקוֹה וְהָיָה יְהִי יְהִי

4Q8

וַיֹּאמֶר אֱלֹהִים יִקְוֹא וְהָיָה מִיֶּסֶד לַשָּׁמַיִם וְהָיָה יְהִי יְהִי
וְיִקְוֹה מִיֶּסֶד הַשָּׁמַיִם לַמְּקוֹה וְהָיָה יְהִי יְהִי

Figure 6: The Hebrew content pool for Gen 1:9 and a few local texts

The content pool is what enables CEDAR to display local texts as if they were layered on top of each other. If we were to layer the Leningrad Codex, 4Q2, and 4Q8 (see the representations of these manuscripts in Figure 6), CEDAR would display agreement in the word *wayyōmer* because the same database items have been used to build that word in all three local texts: the *vav* is the same *vav*; the *yod* is the same *yod*; etc. Because the set of manuscripts where these characters occur has been recorded in the database, alignment algorithms are unnecessary for comparing manuscripts.

Figure 7: A character (*vav*) in the Hebrew content pool and the multiple hierarchies to which it is assigned; notice especially 4Q2, 4Q8, and the Leningrad Codex

¹⁷ Alexandrinus and Coislinianus read γεννητω; MS 509 reads γεννητο.

¹⁸ Incidentally, this strategy is very similar to the collation books used by the Göttingen Septuagint project. Along the left margin of a large ledger page are recorded all attested readings, followed by a list of manuscripts which preserve each reading. Because they were doing it on paper, however, the Göttingen project employed this strategy at the word level; CEDAR employs it at the character level.

This feature of the database model enables the accurate comparison of an unlimited number of manuscripts and eliminates both the duplication of character strings and the vast network of links that would be required by the document model.

Fifth, once local texts have been constructed, they can be linked to digital images of the manuscripts. We can then demarcate areas of the image to link to specific characters or words in the transcription. When a user clicks on the transcription, the linked character on the image is highlighted. Clicking on the image also highlights the linked character in the transcription. This feature, called “hot-spotting,” is particularly useful for studying fragmentary or damaged manuscripts in which reconstructions may be uncertain, because it allows a scholar to communicate precisely which character traces are being interpreted. Furthermore, after a manuscript has been hot-spotted, a user can query the database for all images of a given character. This visual catalog serves as a script chart for evaluating broken characters.

Figure 8: A transcription and image of a Dead Sea Scroll showing a “hot-spot” for the letter *he*; image used by permission of the Israel Antiquities Authority.

One final strategy used by the CEDAR project is that we compare languages by aligning them through the discourse hierarchies of their content pools.¹⁹ Multiple items in one content pool can align with one item in another. The example in Figure 9 shows two cases of multiple Greek words aligned with one Hebrew word, as well as a case of one Greek word aligned with multiple Hebrew words.

Figure 9: Alignment between the Greek and Hebrew content pools in Gen 1:2

Because each local text is built out of the items in its content pool, aligning the content pools allows us to compare any two or more manuscripts in as many languages without duplicating data. We can compare

¹⁹ Recall that in CEDAR texts are organized in two overlapping hierarchies: an epigraphic hierarchy and a discourse hierarchy. Content pools, while strictly speaking not texts themselves, nevertheless are still organized into these same hierarchies. The content pool for each language thus has both an epigraphic and a discourse hierarchy.

any Greek local text with any Hebrew local text – or any several Greek texts with any several Hebrew texts – using only the alignment shown in Figure 9. Any new Greek or Hebrew local texts that we add will also automatically be aligned.

Figure 10: Aligning content pools allows for comparison between any local texts without duplication of data

In cases where the translation is sufficiently loose that it is impossible to establish a one-to-one relationship (or even a one-to-many relationship) between words in the content pools, we have the option to create higher-level structures in the discourse hierarchy. We can create phrases that align with phrases, as in Figure 11, or if need be we can go higher and align entire verses with entire verses, entire paragraphs with entire paragraphs, and so on.

Figure 11: Alignment between phrases when word-to-word alignment is not possible (Gen 1:12)

The Greek, Aramaic, and Syriac content pools are all aligned with the Hebrew because they are translations of the Hebrew.²⁰ When a user asks to compare a Greek local text with a Syriac local text, that relationship is traced through the Hebrew content pool.

Figure 12: Comparing Greek and Syriac local texts through their alignment with the Hebrew content pool

²⁰ At this time, we are focusing our efforts on entering data that pertains to the textual history of the Hebrew. In the future we hope to include data relevant to the textual history of the Greek, such as the Old Latin and Syrohexapla.

Latin, despite also being a translation of the Hebrew, is aligned primarily with the Greek (see Figure 11) because as Indo-European languages Latin and Greek share syntactic and morphological structures which would be obscured if we compared them through the Hebrew. (For example, in the opening words of Genesis, the relationships *in/έν* and *principio/ἀρχή* would be lost if we aligned both languages with the Hebrew בראשית. Instead the relationship would be recorded as *in principio/έν ἀρχή*. While that level of detail would be sufficient for some kinds of research, textual scholars need their data to be as granular as possible.) Thanks to the flexibility of CEDAR, in cases where the Hebrew and Greek diverge and the Latin follows the Hebrew, we can align the relevant Latin discourse units with the Hebrew.

Figure 13: Aligning the Latin with both Greek and Hebrew (Gen 1:7)

5 Using the digital HBCE

The benefit of CEDAR's organized system of reusable content is that rather than reading a list of decontextualized variants in the critical apparatus, users of the HBCE digital edition will have access to all of the primary data used by the scholar in constructing the edition, in addition to the tools to analyze it. Ronald Hendel's edition of Gen 1:9 for the HBCE will serve as a demonstration of CEDAR's potential.²¹

Let us say that we are reading HBCE Genesis 1, and we come to verse 9: וַיֹּאמֶר אֱלֹהִים יִקְוּ הַמַּיִם מִתַּחַת הַשָּׁמַיִם 9: וַיֹּאמֶר אֱלֹהִים יִקְוּ הַמַּיִם מִתַּחַת הַשָּׁמַיִם אֶל־מְקוֹמָם וַתֵּרָא הַיַּבְשָׁה. ²² The accompanying text-critical notes indicate that there are multiple divergences in this verse from the Masoretic Text, and we decide to investigate. We first ask CEDAR to display HBCE as the base text with the Leningrad Codex overlaid in order to see more specifically how HBCE differs from the Masoretic Text. CEDAR uses color-coding to inform the user efficiently about the relationships between the texts.²³ Green indicates that the texts agree. Black indicates characters in the base text (in this case, Hendel's HBCE edition) that are not attested in the overlaid text (in this case, the Leningrad Codex). Red indicates disagreement between texts. When we click on the red *he*, a pop-up window (see Figure 14, bottom left) provides relevant information about that character. The top half of the window shows information related to the texts on display in this particular view. We see that instead of HBCE's *he*, the Leningrad Codex reads *mem*, resulting in a different word: מקום instead of מקוה. The bottom half of the pop-up window shows information about texts in the project that are not included in this view. We can see that the *he* reading is shared by the Dead Sea Scroll 4Q8 and the retroversion of the Septuagint (see left red box). When we click on a black character, say the circled *tav*, a similar window (see Figure 14, bottom right) informs us that while the Leningrad Codex lacks the black text, the *tav* at least is attested in 4Q10 and the retroversion of the Septuagint (see right red box).

²¹ Ronald Hendel's critical edition of Genesis is forthcoming in the HBCE series.

²² "And God said, 'Let the waters below the heavens be gathered to one gathering so that the dry land may appear,' and it was so. The waters below the heavens were gathered to their gatherings, and the dry land appeared."

²³ Our use of color-coding is not necessarily a long-term solution given the obstacle it presents for color-blind users.

The screenshot shows the HBCE Genesis software interface. The main window displays the Leningrad Codex overlaid on the HBCE text. The text is in Hebrew, with English translations provided. The interface includes a top menu bar with options like 'Include MORE', 'Score format', 'Display frequencies', 'Use size to indicate frequency', 'Show allographic variants', 'Show damage', and 'Show scribal phenomena'. Below the menu bar, there are tabs for 'Document View' and 'Analytical View'. The main text area shows the Leningrad Codex text, with the HBCE text overlaid. The text is in Hebrew, with English translations provided. The interface includes a top menu bar with options like 'Include MORE', 'Score format', 'Display frequencies', 'Use size to indicate frequency', 'Show allographic variants', 'Show damage', and 'Show scribal phenomena'. Below the menu bar, there are tabs for 'Document View' and 'Analytical View'. The main text area shows the Leningrad Codex text, with the HBCE text overlaid. The text is in Hebrew, with English translations provided.

Figure 14: HBCE Genesis with Leningrad Codex overlaid; pop-up windows provide details

Figure 14: HBCE Genesis with Leningrad Codex overlaid; pop-up windows provide details

Based on the information from the overlaid view, we now want to look at 4Q8 and 4Q10 along with the Septuagint manuscripts that attest this verse. When comparing across languages, CEDAR displays transcriptions side-by-side rather than overlaid. When we click on a word in any of the transcriptions, CEDAR highlights the equivalent word in all the rest. When we click on מקוה in the HBCE edition, we can see the agreement in 4Q8 and the three Greek manuscripts. Nothing is highlighted in the Leningrad Codex because it has a different reading; nothing is highlighted in 4Q10 because it is fragmentary and does not attest this verse.

The screenshot shows the CEDAR software interface. The main window displays the HBCE Genesis text, with the Leningrad Codex overlaid. The text is in Hebrew, with English translations provided. The interface includes a top menu bar with options like 'Include MORE', 'Score format', 'Display frequencies', 'Use size to indicate frequency', 'Show allographic variants', 'Show damage', and 'Show scribal phenomena'. Below the menu bar, there are tabs for 'Document View' and 'Analytical View'. The main text area shows the Leningrad Codex text, with the HBCE text overlaid. The text is in Hebrew, with English translations provided.

Figure 15: Comparison across languages with equivalent terms highlighted

Figure 15: Comparison across languages with equivalent terms highlighted

Transcriptions are useful, but we want to see the primary evidence, so we ask CEDAR to display the image of 4Q8. The *he* is unequivocal.

Figure 16: High-resolution digital image of a Dead Sea Scroll fragment (4Q8^a) linked to its transcription; image used by permission of the Israel Antiquities Authority

Satisfied with that reading, we move to the Greek manuscripts. A full page of Codex Alexandrinus is time-consuming to navigate, so in the transcription we click on $\text{C}\bar{\Upsilon}\text{A}\Gamma\Omega\Gamma\text{H}\text{N}$, the word we want to see, and CEDAR highlights it on the image.

Figure 17: Digital image of Greek codex (Alexandrinus) with hot-spot. The left pane shows the transcription arranged by line, the center pane shows the transcription arranged by verse, and the right pane shows the image of the manuscript. Clicking in any of the three panes highlights corresponding content in the other two; image of *Facsimile of Codex Alexandrinus* (London: British Museum, 1881), vol. 1, fol. 1r used by permission of Special Collections Research Center, University of Chicago Library.

We repeat this process for the other two Greek manuscripts and confirm that the reading is sound.

Having worked through the evidence for the first non-Masoretic reading in the HBCE edition of Gen 1:9, we turn to the second. When we click on וּתְרִיא , we see agreement in 4Q10 and again in the Greek manuscripts. 4Q10 preserves just part of a much longer plus which is attested in its full form in the Greek witnesses.²⁴

²⁴ On principle, the CEDAR team makes as few editorial decisions as possible so as not to hinder future scholarly analysis. At times, however, we are forced to make decisions, as here when we needed to decide how to relate this very small fragment of 4Q10 to the rest of Genesis: does it attest the end of the Septuagint plus, as we have chosen to encode it, or is it instead an apocopated form of וּתְרִיא earlier in the verse? While our decision was based on philological research (the wəyiqtol of וּתְרִיא is not attested in an apocopated form in MT), another scholar could reach a different conclusion. Because of its ability to encode multiple overlapping hierarchies, CEDAR allows for a future scholar to align the text of 4Q10 differently. This new analysis will not overwrite our initial analysis; instead, both will exist in the database, attributed to the appropriate researcher and available for subsequent use.

<p>HBCE Genesis</p> <p>הַמֵּלִים אֲשֶׁר מִתַּחַת לָרָקִיעַ וּבֵין הַמֵּלִים אֲשֶׁר מַעַל לָרָקִיעַ וְהֵי יְהוָה׃ Gen 1:8</p> <p>וַיִּקְרָא אֱלֹהִים לָרָקִיעַ שָׁמַיִם וְהָיָה עֶרֶב וְהָיָה בֹקֶר יוֹם שֵׁנִי׃ Gen 1:9</p> <p>וַיֹּאמֶר אֱלֹהִים יִקְוֹ הַמַּיִם מִתַּחַת הַשָּׁמַיִם אֶל־מָקוֹם אֶחָד וּמִרְאֵה הַיַּבֵּשָׁה וְהָיָה כֵן וַיִּקְוֹ הַמַּיִם מִתַּחַת הַשָּׁמַיִם אֶל־מָקוֹמָהּ׃ Gen 1:10</p> <p>וַיִּקְרָא אֱלֹהִים לַיַּבֵּשָׁה אֶרֶץ וּלְמִקְוֵה הַמַּיִם יָמִים בְּרָא אֱלֹהִים כִּי־קוֹב׃ Gen 1:11</p> <p>וַיֹּאמֶר אֱלֹהִים תְּדַשֵּׂא הָאָרֶץ דָּשָׂא</p>	<p>4Q8</p> <p>Gen 1:8 — שׁ Gen 1:9 — וַיֹּאמֶר — מְקוֹה Gen 1:10 — לַיַּבֵּשָׁה —</p>	<p>4Q10</p> <p>broken — וְהָיָה broken Gen 1:9 Gen 1:13 — vacat Gen 1:14 — יָמִים יִהְיֶה מֵאֲרוֹת בְּרִיקָע Gen 1:15 — הַשָּׁמַיִם לְהַבְדִּיל — יוֹן הַלֵּילָה וְהָיָה Gen 1:16 — לְאֹתוֹת וּלְמִעֲדִים וּלְיָמִים וְלֵאמֹר Gen 1:17 — עַל הָאָרֶץ וְהָיָה כֵן Gen 1:16 Gen 1:18 — הָאָדָם — Gen 1:19 — קִבָּה בְּרָא — Gen 1:20 — וּמֵל —</p>	<p>Codex Alexandrinus</p> <p>MECON TOY YAAOTOC TOY EPANQ TOY CTEPEOMATOC vacat Gen 1:8 KAI EKALACEN O EC TO CTEPEOMA OYFANON KAI IDEN O EC OTI KALON KAI EGENETO ECITEPA KAI EGENETO PRQT HMERA DEYTERA vacat Gen 1:9 KAI EIPEN O EC CYNAXETHQ TO YAP TO YPOKATQ TOY OYNQY EIC CYNAGHFN MIAN KAI OFETHQ H SHPA KAI EGENETO OYTQC KAI CYNHXQH TO YAP TO YPOKATQ TOY OYNQY EIC TAC CYNAGHFC AYTON KAI OFETHQ H SHPA Gen 1:10 KAI EKALACE O EC THN SHPAN HFN KAI TA CYCTEMATA TON YAAOTQ N'EKALACEN QALACCAC KAI IDEN O EC OTI KALON Gen 1:11 KAI EIPEN O EC BAACTHCATQ H FH BOTANH XOPTQY CTEPON</p>
<p>Codex Bezae Cantabrigiae</p> <p>ANA MECON TOY YAAOTOC TOY EPANQ TOY CTEPEOMATOC Gen 1:8 KAI EKALACEN O EC TO CTEPEOMA OYFANON KAI IDEN O EC OTI KALON KAI EGENETO ECITEPA KAI EGENETO PRQT HMERA DEYTERA Gen 1:9 KAI EIPEN O EC CYNAXETHQ TO YAP TO YPOKATQ TOY OYNQY EIC TAC CYNAGHFN MIAN KAI OFETHQ H SHPA KAI EGENETO OYTQC KAI CYNHXQH TO YAP TO YPOKATQ TOY OYNQY EIC TAC CYNAGHFC AYTON KAI OFETHQ H SHPA Gen 1:10 KAI EKALACE O EC THN SHPAN HFN KAI TA CYCTEMATA TON YAAOTQ N'EKALACEN QALACCAC KAI IDEN O EC OTI KALON vacat Gen 1:11 KAI EIPEN O EC BAACTHCATQ</p>	<p>Leninrad Codex</p> <p>הַמֵּלִים אֲשֶׁר מִתַּחַת לָרָקִיעַ וּבֵין הַמֵּלִים אֲשֶׁר מַעַל לָרָקִיעַ וְהֵי יְהוָה׃ Gen 1:8</p> <p>וַיִּקְרָא אֱלֹהִים לָרָקִיעַ שָׁמַיִם וְהָיָה עֶרֶב וְהָיָה בֹקֶר יוֹם שֵׁנִי׃ Gen 1:9</p> <p>וַיֹּאמֶר אֱלֹהִים יִקְוֹ הַמַּיִם מִתַּחַת הַשָּׁמַיִם אֶל־מָקוֹם אֶחָד וּמִרְאֵה הַיַּבֵּשָׁה וְהָיָה כֵן׃ Gen 1:10</p> <p>וַיִּקְרָא אֱלֹהִים לַיַּבֵּשָׁה אֶרֶץ וּלְמִקְוֵה הַמַּיִם יָמִים בְּרָא אֱלֹהִים כִּי־קוֹב׃ Gen 1:11</p> <p>וַיֹּאמֶר אֱלֹהִים תְּדַשֵּׂא הָאָרֶץ דָּשָׂא עֵשֶׂב מְרִיעַ לְרֹעַ עֵץ פְּרִי עֵשֶׂב פְּרִי לְמִינֵהוּ</p>	<p>Ms 509</p> <p>ANA MECON TOY YAAOTOC TOY EPANQ TOY CTEPEOMATOC Gen 1:8 KAI EKALACEN O EC TO CTEPEOMA OYFANON KAI IDEN O EC OTI KALON KAI EGENETO ECITEPA KAI EGENETO PRQT HMERA DEYTERA Gen 1:9 KAI EIPEN O EC CYNAXETHQ TO YAP TO YPOKATQ TOY OYNQY EIC CYNAGHFN MIAN KAI OFETHQ H SHPA KAI EGENETO OYTQC KAI CYNHXQH TO YAP TO YPOKATQ TOY OYNQY EIC TAC CYNAGHFC AYTON KAI OFETHQ H SHPA Gen 1:10 KAI EKALACE O EC THN SHPAN HFN KAI TA CYCTEMATA TON YAAOTQ N'EKALACEN QALACCAC KAI IDEN O EC OTI KALON Gen 1:11 KAI EIPEN O EC BAACTHCATQ H FH PON CTEPEMA KATA GENOC KAI</p>	<p>Codex Bezae Cantabrigiae</p> <p>ANA MECON TOY YAAOTOC TOY EPANQ TOY CTEPEOMATOC Gen 1:8 KAI EKALACEN O EC TO CTEPEOMA OYFANON KAI IDEN O EC OTI KALON KAI EGENETO ECITEPA KAI EGENETO PRQT HMERA DEYTERA Gen 1:9 KAI EIPEN O EC CYNAXETHQ TO YAP TO YPOKATQ TOY OYNQY EIC CYNAGHFN MIAN KAI OFETHQ H SHPA KAI EGENETO OYTQC KAI CYNHXQH TO YAP TO YPOKATQ TOY OYNQY EIC TAC CYNAGHFC AYTON KAI OFETHQ H SHPA Gen 1:10 KAI EKALACE O EC THN SHPAN HFN KAI TA CYCTEMATA TON YAAOTQ N'EKALACEN QALACCAC KAI IDEN O EC OTI KALON Gen 1:11 KAI EIPEN O EC BAACTHCATQ H FH PON CTEPEMA KATA GENOC KAI</p>

Figure 18: A second comparison across languages

Once again we check the transcriptions against the primary evidence in the images of the manuscripts.

Figure 19: Image of 4Q10 with transcription ; image used by permission of the Israel Antiquities Authority

Figure 20: Codex Alexandrinus with a second hot-spot; image of Facsimile of Codex Alexandrinus (London: British Museum, 1881), vol. 1, fol. 1r used by permission of Special Collections Research Center, University of Chicago Library.

While Figure 20 shows a hot-spot only for the word *οφθη*, HBCE editors can choose which structures they want to highlight. An editor could, for example, choose to make the entire plus a single hot-spot so that clicking anywhere in the text of the plus would highlight the whole. Because every element of a text is a database item in CEDAR – from biblical verses to lines of text to phrases to words to individual characters and diacritical marks – editors have the freedom to hot-spot any feature of a text at any level of detail.

This example demonstrates the value of the HBCE digital edition. Although there are still advantages to the print edition – reading comfort and the ability to access the resource without an internet connection,

to name two – the digital edition can make available a wealth of primary evidence to a degree that print technology cannot match. Furthermore, the evidence is encoded so that it is granular and easily manipulable, creating opportunities for new observations that in turn will lead to new research.

A few distinctive features of the print HBCE which were not demonstrated in this example deserve mention. First, when an editor determines that the textual data represent multiple editions, the editions are printed in parallel columns (this feature is not applicable to the example of Gen 1:9). In the digital HBCE, both editions will be entered into the database as local texts which the user can choose to view in parallel columns or, for a detailed comparison, in an overlaid view (see Figure 14). Second, the print HBCE includes a critical apparatus and commentary that describe the relationships among the variants and, when appropriate, suggest scribal motivations. In the digital HBCE, these notes will be linked to the critical text and will be viewed in a parallel pane. Although this feature has not yet been utilized in CEDAR, it has been used by other OCHRE projects, such as The Ras Shamra Tablet Inventory;²⁵ including it in CEDAR will be relatively straightforward.

Having provided an overview of the digital edition of the HBCE and the CEDAR project, we would like to offer a few methodological thoughts. First, the goal of this project is to support and assist scholarly analysis, not supplant it. We are not teaching computers to do textual criticism. Rather, we want to provide unprecedented amounts of data to researchers in a form that allows them to do their work more efficiently and in greater detail. Those researchers will still need to understand the significance of the data and know how to use it.

Second, we make no claims about the authority or primacy of any given manuscript or textual form. We do not even imply it by the way we organize our data. In CEDAR, all texts are given equal value. This point is important methodologically because we leave it up to scholars to determine, on a case-by-case basis, which reading is original.²⁶ Furthermore, the CEDAR data model allows scholars to formulate their research questions with a great degree of flexibility. The user interface does not insist, for example, that the Leningrad Codex serve as the base text against which all other manuscripts are compared. It is entirely at the user's discretion to decide which of the many manuscripts should serve as the base text and which manuscripts should be added as comparisons.

6 Conclusion

The digital HBCE will expand the utility of the print edition without losing any of its distinctive features. Scholars will retain full access to the critical text, notes, and commentary of the HBCE while gaining all of the primary evidence that the editors used to construct their editions. The digital edition will thus offer a fuller realization of Hendel's goals for the apparatus of the print edition: "that it be clear, complete, and refutable."²⁷ Even the very best apparatus presents only a sliver of the available evidence, both by design, to focus on what the editor has deemed relevant, and by necessity, because an apparatus must be brief if it is to fit on a printed page along with the text it refers to. The presentation of evidence in the digital HBCE, by contrast, is limited only by the time required for the data entry. Furthermore, CEDAR enables kinds of functionality, such as layering texts and hot-spotting images of manuscripts, that are technologically impossible in print. The digital HBCE will both complement and expand the print edition.²⁸

²⁵ See <https://projects.rcc.uchicago.edu/ochre/RSTI/index.html> [accessed 1 April 2019].

²⁶ "In principle, all ancient readings have an equal status, without relation to the text or translation in which they are found. Although there is certainly some statistical validity for the preference of certain textual witnesses over others, this judgment should not influence the evaluation of individual readings. Statistical information is irrelevant when data are evaluated" (Tov, *Textual Criticism*, 272).

²⁷ Hendel, *New Edition*, 37.

²⁸ A version of this paper was presented at the annual meeting of the Society of Biblical Literature on 17 November 2018. A recorded slideshow of that presentation, "A Preliminary Report from the CEDAR Project," can be found at <https://cedar.uchicago.edu/test-cases/>. Our thanks to Ronald Hendel and Jamie Carr for their comments on the paper. We also wish to acknowledge David Schloen, who conceived of the CEDAR project, and Jeffrey Stackert, who has greatly enriched CEDAR with his engagement, support, and enthusiasm. Finally, our thanks to Simeon Chavel, Joseph Cross, and Doren Snoek for their contributions to CEDAR.

References

- Brown-deVost, Bronson. "Scripta Qumranica Electronica (2016–2021)." *Hebrew Bible and Ancient Israel* 5 (2016): 307–15. DOI:10.1628/186870316X14805961757430.
- Buzzetti, Dino. "Digital Representation and the Text Model." *New Literary History* 33 (2002): 61–88.
- Cummings, James. "The Text Encoding Initiative and the Study of Literature." In *Digital Literary Studies*, edited by Ray Siemens and Susan Schreibman, 451–76. Oxford: Blackwell, 2007.
- Hendel, Ronald. *Steps to a New Edition of the Hebrew Bible*. Text-Critical Studies 10. Atlanta: SBL, 2016.
- Schloen, David and Sandra Schloen. "Beyond Gutenberg: Transcending the Document Paradigm in Digital Humanities." *DHQ* 8 (2014). <http://www.digitalhumanities.org/dhq/vol/8/4/000196/000196.html#d18957e294>, accessed 20 December 2018.
- Schloen, David and Sandra Schloen. *OCHRE: An Online Cultural and Historical Research Environment*. Winona Lake: Eisenbrauns, 2012.
- Tov, Emanuel. *Textual Criticism of the Hebrew Bible*. 3rd ed. Minneapolis: Fortress, 2012.
- Yardney, Sarah, Sandra Schloen, and Miller Prosser. "A Preliminary Report from the CEDAR Project." <https://cedar.uchicago.edu/test-cases/>, accessed 20 December 2018.