Abstract
The definition of text is still a live issue with important implications for theorizing, developing, recognizing, and using emerging forms of digital textuality. This article proposes that no single definition of text is sufficient to account for all manifestations of textuality and presents medieval textuality as a test case. At least four different models for text can be offered, corresponding to the ways in which modern medievalists approach medieval texts: facsimiles, transcriptions, editions, and visualizations. This article argues for the value of studying medieval texts: not only to support historically informed theories of reading and writing, but also, at a time when digital texts are still contesting the assumptions and conventions of print, to suggest alternative models of organizing, representing, and processing textual information.

Keywords
Text; Definitions; Modelling; Reading; Medieval texts

Introduction
In 687, Cuthbert, bishop of the monastic community of Lindisfarne in the north of England, died and was buried in the church. In 698, his body was excavated and found to be undecayed, a circumstance taken as a sign of sanctity. The remains were reinterred in an oak coffin that was carved with images of apostles and angels, as well as inscriptions in Roman and runic letters giving the names of the figures. Probably at this time, a small book was also placed in the coffin with the body: a copy of the Gospel

Yin Liu is Associate Professor of English at the University of Saskatchewan. Her areas of research and teaching include medieval literature, English historical linguistics, and digital humanities. This article is a contribution to the Medieval Codes project (http://medievalcodes.ca). Email: yin.liu@usask.ca.
of St John in Latin, now known to scholars as the St Cuthbert Gospel (formerly the Stonyhurst Gospel) – the oldest surviving complete book, with its original binding, in Europe. At a later date in the tenth or eleventh century, the body was clothed in new vestments, including a stole with the names of Baruch and Habakkuk on it, and a silk cloth, probably of Byzantine origin, bearing a now-unreadable inscription in Greek (Rollason & Dobson, 2004; British Library, 2012; Bonner, Rollason, & Stancliffe, 1989; Bonner, 1989). Thus, when members of this medieval saint’s community wished to venerate his body, they surrounded it with texts.

It is surely uncontroversial to claim that the documents I have described – the stole, the so-called “Nature Goddess Silk,” the coffin, and especially the gospel book, contained texts of various sorts. Nevertheless, the ways in which these texts were encoded and used may strike us as bizarre. Cuthbert’s body was invested with at least two instances of wearable text, and on both the stole and the coffin there is an obvious symbiosis of text and image. The texts placed in and on his coffin were encoded in three different alphabets – Greek, Roman, and runic – with the further complication that the coffin inscriptions are in both Roman and runic, in a puzzlingly random distribution. The gospel book is both the most familiar and the strangest of all, for it presents a text in a form that we think of as paradigmatic – writing in a code – but, by being placed in a coffin, it was rendered inaccessible to readers for four centuries, although we often assume that the primary function of a book is to be read. In 875, after Viking raids on Lindisfarne, Cuthbert’s coffin was taken inland to Chester-le-Street, then to Ripon, and then finally to Durham Cathedral. There the gospel book was kept and venerated as a relic. In the twelfth century, Reginald of Durham recorded that those who wished to touch it were required to fast beforehand, and that important visitors were sometimes allowed, as an honour, to wear it, in its little leather bag, around their necks (Tudor, 1989). Thus even this book became a wearable text.

I have used the word “bizarre” to describe one probable reaction to this history of the artifacts surrounding Cuthbert’s dead body, but this little scenario, from the beginning of the history of English literacy, is instructive precisely because it is not really bizarre. It should remind us of two facts about texts. First, we cannot expect that people in other historical and cultural situations (for example, early medieval England) conceptualized and used texts in the same ways that we do. Early medieval English reading practices and processes can be seen as expressions of “transitional literacy” (O’Keeffe, 1990); books before the eleventh century were not primarily intended to be read (Cavallo & Chartier, 1999); and, as late as the fourteenth century, changing uses of texts in England were both causes and symptoms of social upheaval (Green, 1999). Second, even now we create and use texts in many ways that scholars of text often ignore. For example, we commonly surround our living bodies with wearable texts; a garment bearing the names of the rather obscure biblical figures Baruch and Habakkuk may be strange to us, but the concept of a garment bearing names is not. Still, theories of textuality in the English-speaking world in the early twenty-first century tend to be strongly biased toward one particular model of text, in which a text (the indefinite article is significant; see Caton, 2013) is conceptualized as stable, structured, bounded, accessed through a “book,” and intended for reading. This model is not only an historical outcome of print technologies and their social contexts, especially from the eighteenth century onward, but also an ideologically focused
Selection from among many possible models of textuality that have been available since at least the Middle Ages.

Debates about the meaning and nature of text persist and have been sharpened by the anxieties, challenges, and opportunities of realizing textuality in digital forms. Some of these debates exist because it is assumed that there is a singular model of text that must necessarily form the theoretical basis of any attempts to represent and process texts in print or digital environments, and thus we argue about which model of text is the most accurate, functional, reasonable, productive, or rigorous. Such dogged, even dogmatic, essentialism would be scorned in many other areas of the humanities, but its appearance in the digital humanities is surprisingly taken for granted. Thus the provocative article (DeRose, Durand, Mylonas, & Renear, 1990) that presented us with the Ordered Hierarchy of Content Objects (OHCO) model that underlies Standard Generalized Markup Language (SGML) and Extensible Markup Language (XML) was entitled “What is Text, Really?” and it was provocative not because of the question it asked but because of the adverb it attached to the question. Even when, responding to criticism of the OHCO model and its implications for text encoding, Allen Renear qualified and complicated his definition of text, he clung to the ontological claim of really and entitled the sequel “Refining Our Notion of What Text Really Is” (Renear, Mylonas, & Durand, 1993). The debate has been continued by McGann, Eggert, and Robinson, among others (see Biggs & Huitfeldt, 1997; Renear, 1997; Renear & McGann, 1999; Eggert, 2005; Robinson, 2009). But while scholars argue over what text really is, around them are millions of people creating and using texts without committing themselves to any kind of textual metaphysics and negotiating – often with ease and very often unreflectively – multiple, variable, and complex texts and textual functions. Furthermore, people have been doing this for thousands of years. One advantage of approaching text modelling from the perspective of a medievalist is that the very strangeness of the medieval context alerts us to familiarities we might not otherwise have noticed. Another advantage is that the longer historical view helps to prevent us from assuming that any property of text is universal or necessary when it may be a product of our own historical position and cultural prejudices. By illustrating the following discussion with examples from medieval textual situations as well as modern ones, I hope to suggest that consideration even of a dead seventh-century saint may have something to contribute to issues that concern living users of digital text technologies.

If we ought not to insist upon a single model of text, recognizing the diversity not only of text forms but also of approaches to text and uses for text, value remains in attempting to describe possible models of text and their implications. As an invitation to further thought, not as a final word on these matters, I present here four possible models of text that a medievalist might encounter, both with respect to medieval texts and with respect to modern attempts to remediate them. In each case, I outline a provisional definition of text – what we model; an expression of text – how we model it; and a function of text – why we model it. I emphasize from the outset that four models do not exhaust the possibilities; as well, the terminology I suggest for these models is provisional, and I would welcome suggestions for more precise (or at least more eloquent) alternatives.
**Material text**

**Definition:** Text is a sequence of symbols expressed in some material form

A *material* model of text privileges the text’s historical existence in the physical world. The text exists in some kind of *medium*: stone, bone, wood, cloth, parchment, paper, or digital environment. (Although I use the present tense, I include texts that once had a physical existence but no longer do, such as “lost” manuscripts.) For the most part, we are accustomed to think of texts as being expressed by physical entities called *documents* (Shillingsburg, 1996; Brown & Duguid, 2000), paradigmatically in the media of parchment or paper, but now in digital files as well. However, texts can also be inscribed on coins, stone memorials, household objects, and the like. Thus not only is the gospel book from Cuthbert’s coffin a document; so are some of the vestments in which his body was clothed and the inscribed coffin itself.

**Expression:** Facsimile

Facsimiles are natural expressions of a material model of text. They are most often visual and two-dimensional representations of documents; however, I wish the term *facsimile* here to include other sensory data (e.g., tactile), or three-dimensional facsimiles of artifacts. Thus the digital facsimile of the St Cuthbert Gospel, available on the British Library website, provides visual images of the book. However, because a physical document is unique and irreproducible, all facsimiles are approximations of the documents they represent. No matter how high the resolution may be for visual images of the St Cuthbert Gospel, no visual image can capture the tactile experience of holding the book, nor will most photographic facsimiles show features that are visually difficult to detect, such as drypoint markings on parchment or watermarks on paper.

**Function:** Viewing

Facsimiles provide ways of viewing text in its document context (see Echard, 2008, pp. 208–209). What is important in this approach to text is not so much what the text says, but how it looks: its visual appearance and spatial location, usually, and its relation to its material context and medium. It is safe to say that the British Library did not digitize the St Cuthbert Gospel primarily so that people could read the Latin text of the Gospel of St John – which, after all, is readily available in many other forms. When the British Library acquired the manuscript in 2012 and created a special exhibition for it, most visitors who came to look at the book probably could not read Latin nor did they wish to; they came to see the material object and to learn about its history. Similarly, the twelfth-century monastic community that preserved Cuthbert’s relics at Durham seem to have had little or no interest in reading the book; it was important to them primarily as a symbolic object whose material form was an expression of spiritual values and powers. Thus “viewing” may not be a simple visual encounter with a document; it may be charged with emotional significance and may extend to handling the physical object or incorporating it into particular social practices. When the British Library created its “Turning the Pages” animation for the digitized books it featured on its website, it deliberately tried to reproduce not only the images of the pages themselves but also the action of leafing through them, although the animation arguably adds nothing to the reading of the text itself.

For a modern scholar, viewing a facsimile (or, better, working with the document itself) offers information that cannot be obtained in other ways. A codicologist might note that the binding of the book uses a method known as “Coptic sewing,” which was much more common in the Mediterranean region than in northwestern Europe (Brown, 2007). A paleographer looking at a page of the St Cuthbert Gospel will notice that the uncial script declares the document’s allegiance to a text tradition developed in Rome in the fourth century, and consciously followed in Bibles produced in the monasteries of Monkwearmouth and Jarrow under Abbot Ceolfrith in the late seventh and early eighth centuries. Nevertheless, the clear word separation in this script is an insular innovation, developed earliest in Irish writing of Latin texts (Roberts, 2005, p.13; Brown & Lovett, 1999, p. 41–44; Saenger, 1997, p. 81–87). It may be a point of debate as to whether this sort of information, gained by viewing, is properly textual or might be described as “contextual” instead, but it nevertheless depends on the ways in which the text is presented and encoded in the physical document. I suspect that most people today would regard word-bounding space as an essential component of text, but it was a new encoding element in the seventh century.

**Definition: Text is a set of abstract linguistic structures**

A structural model of text has perhaps its most influential description in the so-called OHCO thesis, which defines text as “an ordered hierarchy of content objects,” and serves as the model on which SGML and its special case, XML, are based (DeRose et al., 1990; Renear et al., 1993). Allen Renear has usefully laid out the philosophical basis of this definition as expressed in five statements – that texts are

- real: they have properties independent of our interests in them and our theories about them.
- abstract: the objects which constitute texts are abstract, not material, objects.
- intentional: texts are, necessarily, the product of mental acts.
- hierarchical: the structure of texts is fundamentally hierarchical.
- linguistic: texts are linguistic objects; renditional features are not parts of texts, and therefore not proper locations for textual meaning (Renear & McGann, 1999).
The point that textual structures are necessarily hierarchical is hotly debated, but it is not, as I take it, essential to a structural model of text. (Thus a nonhierarchical structural definition of text would correspond to the theory that Renear calls “pluralist”; see Renear, 1997). Textual structures may or may not be hierarchical, but they are also not necessarily sequential.

The crucial point of difference between a structural and a material model of text is that a structural model presents text as an abstraction, as distinct from its material expressions. Unlike a material text, a structural text is reproducible, and indeed one of its tests – and its functions – is its reproducibility. In this model, these words in the same order but a different font would, under normal conditions, constitute the same text. “Renditional features” such as script or font are not considered features of text. This reproducibility means that text can preserve its identity regardless of medium. A structural model of text does not deny that information may be conveyed, or extracted, from the material and/or social environment of the text, but argues that such information would be contextual, not textual. Thus, according to this model, there is nothing remarkable about the text of the St Cuthbert Gospel, which is the Latin Vulgate version of the Gospel of St John; the “same text,” as long as it reproduced the same linguistic structures in the same order, would be found in a print or digital surrogate where the letterforms are modern rather than seventh-century uncial.

**Expression: Transcription**

Transcriptions express a structural model of text. A medieval scribe copying a text was not producing a facsimile but a transcription; to do so, the scribe needed constantly to distinguish between the text being copied and other features to be excluded from the copied text. Because of these decisions, transcription is not simply “data migration”; it amounts to a definition and thus an interpretation of the text being transcribed. The goal of transcription “is not to represent as correctly as possible the originals, but rather to prepare from the original text another text so as to serve as accurately as possible certain interests in the text” (Pichler, 1995, p. 691). Certain features are retained and thus transmitted; others are excluded as extraneous, extratextual, or nonfunctional. Transcription thus, like any model of text, distinguishes (not always easily) between “information” and “noise.”

Cuthbert’s successor at Lindisfarne, Eadfrith, made a much more famous book c. 700, the Lindisfarne Gospels. Although the opening words of the Gospel of St John look strikingly different in Eadfrith’s document than in the book that was put in Cuthbert’s coffin, a seventh-century monk or a twenty-first-century reader might argue that they constitute the same text: *In principio erat verbum…*

**Function: Storing/Transmitting**

The primary purpose of encoding or copying a text in the Middle Ages was to ensure its continuity in time. In today’s world of “information and communication technologies,” information is meant to be retrieved, transferred, and processed. In the 1948 paper that founded information science, Claude Shannon defined information in such a way as to make storage a special case of transmission (Shannon, 1948; Gleick, 2011). But we cannot assume that communication was the primary purpose of a
medieval document. In the monasteries where most book production in early medieval Europe took place, books were aids to spiritual meditation, rather than repositories of information; not until later in the Middle Ages (from about the late eleventh century onward) did the idea of reading a book for its information content become increasingly important (Cavallo & Chartier, 1999). The St Cuthbert Gospel was used and valued, but for most of its history it seems to have been rarely, if ever, read. Copying preserved a text; it did not necessarily “publish” or circulate it.

**Semantic text**

**Definition:** Text is what a reader finds meaningful

A semantic model of text asserts that “a text exists only because a reader gives it meaning” (Cavallo & Chartier, 1999, p. 1). In this model, text is constructed by a reader from the symbols and structures in, or expressed by, the document. Text by this definition is constrained both by the material presence of the document and the linguistic structures that the reader encounters, as well as by whatever communities of practice the reader belongs to. But it also admits ambiguity and variation in interpretation. Thus two different readers can process the same sequence of symbols in two different ways, constructing two different texts. As we might expect, the strongest proponents of this model of text have tended to come from literary perspectives, where ambiguity and interpretative freedom are often highly valued. Text by this definition is reproducible, but in a social rather than a mechanical context. That is, I reproduce a text by persuading others to read it the same way, to find in it the same meanings. Thus every presentation of a text is implicitly an argument, and any text is always dynamic, in that it is always open to contestation and reinterpretation.

A semantic model of text may seem impractically subjective and untestable, but it need not be. Like the material model of text, a semantic model situates the text in a historically specific context; thus whether or not a particular mark on a page (e.g., punctuation) counts as textual information might depend on the reading practices of a culture at that historical moment. However, unlike material and structural models of text, a semantic model implies that text is always unstable. A text has no independent existence; it comes into being when we read it.
Most of the discussion about models of text has focused on the scholarly edition – understandably, given the importance of that genre to academic work. Because this ground has been so well-trodden (some might say, trodden into mire), I will not revisit that discussion in detail here. Suffice it to say that debates over editorial theory are often important to a medievalist, insofar as editions are always recognizably interventions between the document and the reader. For an editor, a document such as the St Cuthbert Gospel becomes a witness, on the basis of which the editor constructs (or perhaps confects) a text. If a transcription is an abstraction from the physical marks in a document, an edition is an abstraction from the transcription(s) available to the editor. An edition both invites and provides a reading of the text, in different senses of the word reading. The reader of an edition is given the information the editor judges necessary to make the text meaningful, but, at the same time, the editor's decisions about how the text is represented constrain and influence possible interpretations.

It is often remarked that a medieval scribe copying a text acted in many ways like a modern editor: he felt free (or perhaps even obligated) to make a text more "readable," whether that meant correcting perceived errors, improving on the exemplar's wording, rearranging material, adding visual elements, or the like. The degree to which a scribe "edited" a text depended not only on scribal idiosyncrasy but also on the type of text being copied; a scribe was not as free to alter the text of Scripture as he was to alter the text of a secular poem in a vernacular language. Celebrations of textual variation in medieval manuscript cultures, coming from a modern culture that values variation as a positive sign of creativity, sometimes neglect the dominant conservatism of medieval mindsets. Reading was primarily oral and public, not silent and private; the invocation of authority was valued, not the attempt at originality. Thus the semantic model of text was enacted somewhat differently in the Middle Ages than it is today. In the St Cuthbert Gospel are four small late-seventh-century annotations in the margins, the phrases "pro defunctis" or "de mortuorum," indicating that the book was used to celebrate the Offices for the Dead: they occur on fols. 20v, 27r, 28v, and 51r, marking respectively John 5:21, 6:37, 6:51, and 11:21. Read in this liturgical context, the text of the Gospel of St John would have acquired particular meanings for its listeners, but those meanings would have been guided by the authority of the Church.

In modern print cultures, the reading function of text, for reasons social and cultural as much as technological, has become, at least ostensibly, the dominant one – thus the disproportionate attention paid to scholarly editing when academic humanists discuss textuality. The dominant image of reading in elite print culture is of an individual thoughtfully perusing a book, beginning on page 1 and continuing to the end without being distracted by the neighbours' children and without cheating by flipping to the end to find out who done it. Early modern printers, discouraged by their technology from efficiently replicating many of the complex textual devices available to medieval scribes, turned the loss of information into a virtue (see Carlson, 2004); the clean, linear typography of the printed page invited a kind of reading to which Western literate elites became, and still are, emotionally attached. But reading is not the only function of a text, even now; and in other historical, social, and cultural contexts, it might not be the primary function at all.
**Text as data**

**Definition: Text is a form of data**

In calling this a data model of text, I am obviously pointing toward computing environments, where text is only one kind of data, alongside numbers, images, and other forms of information. In this model, reading is only one kind of data processing, and humans are not the only processing agents; machines, such as computers, can process text as well. This model therefore challenges assumptions that text is meant to be readable, at least by humans; there are ways of storing, transmitting, analyzing, and processing text that are not intended for human reading.

This model of text is a much older and more basic one than might at first appear. Writing seems to have originated in most cases as an accounting system rather than as a communication system, just as mechanical computers were originally primarily numerical calculators. Both tools extended their functions significantly when people started using them not only for computing but also for communicating (Hobart & Schiffman, 1998).

**Expression: Visualization**

Perhaps the most radical and potentially transformative way in which digital environments can change our approaches to textuality is in the increasing possibilities for visualizing texts rather than reading them. Visualizations express a data model of text.

But there are important medieval precedents for digital visualization as well. Every system for presenting textual information in a form other than the linear sequence in which the symbols would be read aloud amounts to a visualization: marginal commentary, interlinear glosses, canon tables, columnar layout. The St Cuthbert Gospel does not present us with any examples of such data structures, but the Lindisfarne Gospels do. For example, the biblical text in the Lindisfarne Gospels is prefaced by a full set of canon tables, a device developed by Eusebius of Caesarea in the early fourth century to correlate parallel passages in the four canonical gospels (see Brown, 2003, pp. 179–187). Canon tables work by dividing the text into sections (not identical to modern biblical chapter divisions) and representing them as parallel structures in tabular format. In the Lindisfarne Gospels,
the full set of canon tables attempts to organize the text of the gospels in much the same way as a relational database organizes information.

Most of the text of the Lindisfarne Gospels was also provided with interlinear Old English glosses by a priest named Aldred in the tenth century – “with the help of God and St Cuthbert,” he says in the colophon. Glosses of this sort provide a word-for-word translation of the Latin text, creating a one-to-one correspondence between Latin and English texts that is two-dimensional and visual, rather than linear and oral.

FUNCTION: PROCESSING
It is noteworthy that many medieval innovations in the presentation of textual information arose in situations where the written text was not simply a transcript for speech (Parkes, 1999). For example, word separation developed in the British Isles in the seventh and eighth centuries precisely because, for Irish and Anglo-Saxon scribes, Latin was a language associated primarily with writing and only secondarily or parasitically with speech (Saenger, 1997; Wright, 2002). The earliest extensive examples of English writing that we have are glosses on Latin texts. The medieval interlinear gloss, and its associated genre, the glossary, was a pervasive and important medieval data structure, although surprisingly understudied (Healey, 1997). Other forms of reference reading were closely associated with the rise of scholastic culture in the Middle Ages, particularly from the twelfth century onwards. Texts became datasets from which users (e.g., preachers) could select short items of information without reading everything; tools such as indexes, concordances, and florilegia were developed to support this kind of information retrieval (Hamesse, 1999; Rouse & Rouse, 1991). A great number of tools, devices, and strategies for searching and referencing textual information, many of which we still use today, originated or were extensively developed in the Middle Ages.

### Table 1: Four models of text

<table>
<thead>
<tr>
<th>Model</th>
<th>Definition</th>
<th>Expression</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>Text is a sequence of symbols expressed in some material form.</td>
<td>Facsimile</td>
<td>Viewing</td>
</tr>
<tr>
<td>Structural</td>
<td>Text is a set of abstract linguistic structures.</td>
<td>Transcription</td>
<td>Storing/ Transmitting</td>
</tr>
<tr>
<td>Semantic</td>
<td>Text is what a reader finds meaningful.</td>
<td>Edition</td>
<td>Reading.</td>
</tr>
<tr>
<td>Data</td>
<td>Text is a form of data.</td>
<td>Visualization</td>
<td>Processing</td>
</tr>
</tbody>
</table>
I do not have space in this article to discuss further important aspects of medieval textuality. Translation, for example, is a deeply significant dimension of practically any medieval text, not least the St Cuthbert Gospel, which is a Latin translation and which, with Cuthbert’s body, was further translated (in the original sense of “moved from one location to another”) from Lindisfarne to Durham, and eventually to the British Library. It should be noted as well that the four models of textuality outlined here are not to be thought of as watertight compartments; transcriptions and editions, for example, should be considered as points on a continuum rather than as absolute categories. Let these ideas be taken as suggestions toward inquiries and explorations that do not seek to limit the range of possible text models, but to broaden it.

We are at a propitious time when digital technologies are raising questions, not only among scholars but also in the general public, about what texts are, what forms they may take, and how they are used. In seeking answers to such questions, one of our greatest resources is history, for understanding both how modern text technologies originated and developed, and also why they were developed, and why past tools and devices persisted, changed, or became obsolete. Much of this history has been surprisingly neglected – by digital humanists, scholars in media studies, even book historians – partly because it has been traditionally the territory of such esoteric academic creatures as paleographers and codicologists. But, just as number theory was thought to be among the most abstract and unpragmatic branches of mathematics until it was applied to cryptography, epigraphy and manuscript studies ought now to be active contributors to any historically informed study of information technologies. The acquisition of the St Cuthbert Gospel was achieved after the most extensive fundraising campaign in the British Library’s history; the 2013 public exhibition of the Lindisfarne Gospels in Durham has been, by all accounts, a great success. But the opportunity now exists for the study of past texts to move beyond nostalgia, and for dead people and old books to inform and enrich living readers and their new technologies.

Acknowledgements
The ideas in this article come out of research underway for the Medieval Codes project; more information can be found at http://medievalcodes.ca. I gratefully acknowledge the support of the University of Saskatchewan, and especially the Humanities and Fine Arts Digital Research Centre, for the project and for this article in particular.

References


