Collaboration and Dissent: Challenges of Collaborative Standards for Digital Humanities (pre-print version)

Julia Flanders

Collaboration—literally a shared work—is always understood to carry with it some kind of sacrifice, a tradeoff between autonomy and synergy. In our collaborative relationships, we intensify the concessions we make to the demands of the social contract, and we voluntarily submit to norms of behavior and constraints on our freedom of action in order to gain the benefits of a group undertaking: a barn-raising, a collection of essays, a successful conference. But even before we as collaborators can adopt these norms, we (in a larger sense) also have to develop them. The collaboration of conversation is predicated on the norms of language; collaboration on a scholarly edition is predicated on a set of private agreements about the editorial mission, set against the backdrop of larger disciplinary expectations concerning what editing means and how it proceeds. So we might start by observing that collaboration takes place within frames of expectation that may be private, local, professional, or broadly social. The vectors of agreement and conformance are thus not solely between the collaborators themselves, but also between the collaborators and some standards that operate beyond their own sphere of activity.

What makes collaboration between projects in the digital humanities particularly interesting —and particularly challenging—is the way these different frames of expectation intersect. Digital humanities projects take place, strikingly, in a universe constrained by a set of technical norms that govern the informational and operational behavior of the digital environment. Because these collaborations are so often aimed at building something that works—a tool, a resource, an online collection—the collaborative activities are typically mediated through things like software tools and data standards that are quite uncompromising. For instance, XML documents must obey the rules of structural well-formedness prescribed by the XML standard; dates entered into a metadata record must follow the agreed-upon format or the sorting routines that operate on this data won't work. (Geoffrey Nunberg's 2009 critique of Google Books' metadata illustrates the consequences of failure vividly.) These technical standards mediate the collaboration by eliminating the need to create all of these conventions from scratch, and the more exacting they are (and the more exactly they are observed) the more powerful the technical outcome. High-function digital projects arise from strong adherence to the right standards.

But in addition, as these projects arise from humanities research, they also require agreement concerning disciplinary norms that shape the practices of digital representation. These include, for instance, acceptable practices of transcription, regularization and emendation; acceptable standards of authenticity and verification; the kinds of commentary and contextualization that are acceptable or required; and beliefs about the interpretive or analytical or critical goals that are at stake. These norms arise from detailed, ongoing debates concerning both the ultimate goals of scholarship and the methods and practices by which we achieve them.

A traditional humanities view of this picture might ask what these two sets of norms have to do with one another, any more than the editorial standards for a critical edition have to do with the

standards guiding the machinery that binds the book. But what the field of digital humanities as a whole has revealed is precisely how tightly interwoven and mutually consequential "technical" and "disciplinary" standards often are. Our concern in this discussion is in fact a domain in which disciplinary and technical norms overlap and where we see, locked in struggle, the drive towards absolute consistency and technical processability on the one hand, and the drive towards critical independence and disciplinary debate on the other. I am referring here to the arena of standards for digital representation of research materials, and in particular the domain of scholarly markup languages. Standards of this kind are an essential precondition for the development of collaborative projects in the digital humanities, because they permit us to exchange scholarly information about research materials in a digital form.

A markup language is a method of digital representation through which we can create highly complex models of textual artifacts. The markup constitutes an information structure which formalizes, labels, and enhances the source material: it makes explicit the nature and function of each piece of information, and through that explicitness makes it possible for non-human processes (such as computers) to "understand" and make use of the resulting digital object in ways that take advantage of its self-knowledge about its own structure and content. A manuscript letter marked in this way can carry with it the knowledge of which words have been deleted or replaced by the author, and a software system can use that knowledge (for instance, to generate a final edited version of the text). A collection of such manuscripts, if marked consistently, can yield a systematic analysis of their authors' revision practices.

Text markup of this kind, as practiced in the digital humanities world, sits at the juncture of humanities scholarship—textually nuanced, exploratory, and introspective—and digital technology, with its emphasis on formalism, consistency, and upward scalability. As a result its norms carry a double weight: they must achieve some kind of technically actionable uniformity, but they must also express useful scholarly concepts and differentiations. They draw their functional precision and power from the domain of technical norms, and also their capacity to broker collaborative exchange at a practical level. But they draw their relevance and intellectual usefulness from the domain of disciplinary norms, which as we have seen are fundamentally and on principle resistant to standardization, and in fact might be said to thrive like yeast in a state of fermentation (though without drowning in their own byproducts).

Debates about how to balance these competing concerns in the markup community—and particularly in the community of the Text Encoding Initiative—are lively and sustained. The goals of standardization and facilitated interchange were articulated by the planners of the TEI at the earliest stages of their discussion in the Poughkeepsie Planning Conference of 1987, as documented in the first of the so-called "Poughkeepsie Principles":

> Poughkeepsie Principle 1: The guidelines are intended to provide a standard format for data interchange in humanities research. (TEI, 1988)

These goals have remained strongly in view during the succeeding years, during which the TEI Guidelines have undergone sustained development and use and have become one of the most deeply researched features of the digital humanities landscape. At the same time, the challenges of maintaining the double force noted above—the power of both strong formalism and expressive nuance—have become much more vividly apparent. For some developers, the attempt to maintain

expressiveness simply undermines the goal of interchange and places the entire enterprise at risk. Characteristic of this view is the following quote from a 1997 posting by Mark Olsen to the Humanist discussion list, in which he is complaining about the laxity and permissiveness of the TEI Guidelines as a technical standard, a few years after the first public release of those Guidelines:

> The real test of an INTERCHANGE format, however, is … that the format can be automatically converted TO and FROM any number of systems with a minimum of effort. My principal objection to TEI is that it is by far the most difficult representation to convert into something else, because of its expressive power. The more tightly constrained a specification, the easier it is to write converters. It is a BALANCING act, which I do not believe the TEI community has -- because of its make-up and structure -- really tried to perform. (Olsen 1997, emphasis in original)

This position has also been strongly argued in subsequent debates about the nature and value of what is termed "TEI conformance": that is, the formal methods by which adherence to the TEI Guidelines can be assessed. The rationale for the current definition of conformance, in which conforming documents must be a strict subset of the unmodified TEI,[1] focuses on several practical goals involving the exchange of data:

- interchange or integration of documents amongst different researchers or users;
- software specifications for TEI-aware processing tools;
- agreements for the deposit of texts in, and distribution of texts from, archives;
- specifying the form of documents to be produced by or for a given project.

(TEI 2007, 23.3)

The countering position, equally strongly felt and argued, is that one can only achieve such strong standardization and uniformity by eliminating a degree of intellectual depth and nuance that is essential to the intellectual quality of the enterprise. Michael Sperberg-McQueen, one of the original editors of the TEI Guidelines, responded to an earlier version of Mark Olsen's plea above by offering him an extremely reduced and simplified subset of the TEI and

> demonstrating, by a reductio ad absurdum, how reducing a tag set to this size … forces one to omit too much material which can be useful in the encoding of virtually any text, and which is absolutely essential for dealing rationally with some texts. (Sperberg-McQueen 1995)

Sperberg-McQueen's point about textual complexity is even more true when we consider the role markup plays in representing not just texts but our views about texts: our methodological assumptions, our editorial decisions, our critical debates. The prospects for representing these in a

---

[1] That is, they must use only elements and element structures that are defined in the unmodified TEI; any conforming TEI-encoded document must be valid against the unmodified TEI schema or be transformable into a valid document with no loss of information.

simple and uniform manner are generally acknowledged on both sides of the debate to be extremely poor. Clearly these arguments would not apply to projects whose goals for text representation are quite simple (for instance, digital library projects where the transcription serves only to provide searchable full text to accompany a page image). But for projects in which the TEI functions as a representation of scholarly work, the accommodation of nuanced intellectual expression is clearly crucial.

In debates about this issue, it is usually assumed (as Mark Olsen assumes) that a strongly enforced uniformity will facilitate collaboration, and conversely that heterogeneity and dissent will militate against it. Our instinct may thus be, in the context of digital standards, to treat disciplinary debate as the opponent: something that needs to be eliminated or ignored in order for collaboration to proceed. I would like instead to suggest otherwise: that in fact the humanities dimension of digital humanities work makes this elimination impossible and also undesirable. But it will help if we can first establish a more nuanced and carefully structured view of the precise role of dissent within our collaborative ecology.

Collaboration and Dissent

One place where collaboration has been examined from a philosophical and political perspective (rather than a business-efficiency perspective) is in theories of collaborative learning, and John Trimbur in a 1989 essay entitled "Consensus and Difference in Collaborative Learning" provides a particularly thoughtful and critical analysis of some key points: he is considering the dynamics of consensus within learning environments but also making a larger point about the way that social negotiations take place. Trimbur marks consensus as central to collaboration, but for him, a crucial element of the discursive practice of consensus is dissent: not just the need to democratically acknowledge a minority view, but the need to base consensus on "collective explanations of how people differ" (610). This "dissensus" is a representation of fractures within the discourse: an acknowledgement of the existence of the periphery, the discourses out of power:

> We will need, that is, to look at collaborative learning not merely as a process of consensus-making but more important[ly] as a process of identifying differences and locating these differences in relation to each other. The consensus that we ask students to reach in the collaborative classroom will be based not so much on collective agreements as on collective explanations of how people differ, where their differences come from, and whether they can live and work together with these differences. (610).

Dissensus thus functions as a way of establishing a critical relationship between the consensus and the dissenting voices. In other words, dissent occupies an informationally important and illuminating position in the universe: "not as the goal of the conversation but rather as a critical measure to help students identify the structures of power." With this understanding of the role of dissent, Trimbur argues, we can finally see consensus not as a rational, realizable goal (akin to finding out that we really all do want the same thing) but rather as a recognition of the "inexhaustibility of difference" and a will to "organize the conditions in which we live and work accordingly" (615).

This argument offers a very interesting perspective on standards development, and one that understandably doesn't get much play in that domain. However, the digital humanities is in a position to attend to the fact that standards do arise from and represent power structures; they represent the functional homogenization that is one outcome of "community": the consent to being homogenized. And as a community-driven standard arising from the digital humanities, the TEI is rare (but perhaps also characteristic) in taking seriously the legitimacy of dissenting views while also seeking a technically functional outcome. It needs such views to exist—indeed, it relies on their existence as the driving force behind its own onward progress—but it wants them to exist *in relation to the community at large*, as part of the discourse rather than apart from it.

We can now usefully circle back to comment on the word "standard" in the context of Trimbur's observations, especially since in an important sense the TEI is not in fact a standard according to a strict definition of the term. The sphere in which strict definitions operate with relevance to the digital humanities is that of governmental and para-governmental standards (such as the American National Standards Institute), international non-governmental organizations (such as the International Organization for Standards), and industrial standards bodies (such as the World Wide Web Consortium). Standards promulgated by such bodies have the force, if not always of law, at least of definitional hegemony: a standard for measuring photographic film speed or tuning frequencies sets the parameters within which all activities will effectively operate.

Standards by their nature thus impose a uniformity that, in effect, creates consensus by fiat. Once they are in place, we obey them because we value the consensus they represent (and the cohesion and practical efficiency they enable) more than our own individual right to do things differently. Furthermore, within the domain of digital humanities, adherence to (open) standards is framed as a kind of good citizenship, the necessary precondition for a free interchange of data. The world without standards is a world of chaos—a post-Babel cacophony of conflicting and mutually unintelligible voices. But it is also a world of plurality, and the transition from that plurality to the uniformity of the standard is worth noting as an important change of state. The web site of the British Standards Institution includes this succinct statement on the nature of consensus standards, which places its emphasis tellingly on this exact point:

> All formal standards are developed with a period of public enquiry and full consultation. They incorporate the views and expertise of a very wide range of interests from consumers, academia, special interest groups, government, business and industry. As a result, standards represent a consensus on current best practice. (BSI 2010)

The deftness with which we move from that "wide range of interests" (which has a double force: both a variety of differing interests, and a representative set of interests) to the "consensus on current best practice" elides what must necessarily have happened: the elimination or bypassing by one means or another of that original plurality of views.

In fact, hard-headedly, standards present us with a balance or tradeoff: they represent an operation of power, which may have been exercised with more or less attention to minority opinions, and may place more or less constraint on our meaningful freedom of action, and may provide more or less practical benefit to us and to society. Different standards bodies and different standards sit at different points along these three axes. It is hard to feel very put out about not being consulted

about the standard for measuring the twist of single-ply yarn, but many of us may feel a greater stake in the ISO standard for representing the sex of humans, in which the permitted values are 0 (not known), 1 (male), 2 (female), and 9 (not specified) (ISO 2004). But even the best arrangement of these factors—the most open process, the least constraint, the greatest practical benefit—in the end does represent as "consensus" an end-product about which there actually has been, and may still be, disagreement. There is no need to standardize what no one differs on.

Collaboration and the TEI

The TEI occupies a distinctive position within this landscape—first of all, because it does not claim the status of a standard for its Guidelines. They "make recommendations" about methods of representing textual sources in digital form, but these recommendations have their force solely within the sphere of activity of the TEI itself: in order to claim to be using the TEI Guidelines, you have to use the TEI Guidelines. This internal regulation of the TEI community positions the TEI as a standards body *for that community of usage*, and for this reason the TEI is often described as a "community standard," both in its own self-description and by others. But this terminology deserves some scrutiny because, even thus qualified, it aligns the TEI's guiding enterprise with goals of uniformity and consensus that, as we have seen above, have proven problematic from the outset. In order to understand how this "community standard" operates *as a standard* within the TEI community and hence as a collaborative tool, we need to understand the role that dissensus, in Trimbur's sense, may play in its formation and use.

Encoding standards like the TEI are foundationally collaborative technologies: they presume the need and the desire to coordinate shared work, to generalize individual insight to a community, and to support the extension, critique, and reuse of ideas and techniques. But how, concretely, does the TEI figure in a typical digital humanities project? A simple example would be a pair of scholars who are working together on a digital edition of a very long manuscript that must be transcribed and annotated: for instance the Almanacks of Mary Moody Emerson—a 1000-page manuscript currently being edited and encoded by Sandra Petrulionis and Noelle Baker in collaboration with the Women Writers Project (WWP) at Brown University. If they divide up the manuscript, each taking a section, their agreement on a shared encoding system such as the TEI makes it possible for them to work together while also working separately: the encoding standard diminishes the need for day-to-day consultation on details of digital representation. When they combine their work, it should form a consistent whole, and when they hand it at last to the WWP for publication, it should fit in with the other materials in our collection that have also been prepared in this way.

In this typical kind of collaboration, the parties to the shared effort know of each other, and their intention to work together is explicit; in effect, a collaborative vector is established between them that operates in real time, mediated by their use of the encoding standard. As a more complex example, however, take a project like the Electronic Archive of Greek and Latin Epigraphy, which is a federation of epigraphic databases: online collections of ancient inscriptions on stone, metal, clay, and other durable surfaces. Scholars in several different countries want to be able to contribute to such collections, knowing that in doing so they are contributing to a single, vast, comprehensive resource rather than to one of many that are small, isolated, and partial. In order to accomplish this goal, they need a common standard for transcription and editing, and in fact this group is now using the TEI as the basis for their digitization efforts: the TEI-based epigraphic

transcription standard is known as EpiDoc. In this collaboration, although the scholars involved know that they are contributing to a shared resource, they don't necessarily feel themselves to be working at all times directly *with* all of the other scholars: the model is more of a hub and spokes rather than a direct person-to-person network.

This kind of indirect collaboration leads us in turn to a third case which has particular importance for the digital humanities. Imagine the same hub-and-spokes model, but now let us think of the spokes as extending not only into space but into time as well. The recent rise of interest in humanities data curation demonstrates how crucial these longitudinal collaborative vectors can be. Digital resources that we create today do not exhaust their value in the first few years of use; on the contrary, unlike research in the sciences which has a fairly deterministic half-life, humanities research materials have a horizon of usefulness that is much more complex and unpredictable and prolonged. The British Women Romantic Poets project is a substantial collection of Romantic-era poetry by women created at University of California, Davis in the late 1990s; it has been available on the web in its original form for years.[2] But in addition, this project contributed texts from their collection to the WWP's online collection, Women Writers Online, and the BWRP collection is now also included in NINES, the large federated collection of resources for the study of nineteenth-century literature.[3] The WWP and NINES are taking advantage of the intelligibility of the BWRP data—its use of standards like the TEI—to do things with these materials that were not specifically envisioned by their creators. But the intelligibility of the data in itself—the intention that it be reused—constitutes an important collaborative gesture. The "collaboration" here is real and important and yet takes place across time and without direct interaction.

This kind of indirect collaboration is at present rare—and indeed, there is a strong anecdotal "literature" attesting to the difficulties in accomplishing it—but it exercises a powerful mythic influence on the ways we think about how we create digital humanities resources and why. As *digital* humanists we are committed to encoding our data in a standard way so that it can be shared, so that it will remain comprehensible: in fulfillment of an implicit contract with unknown scholars of the future who need to know what we know and understand what we have done. But at the same time, as we've already seen, as digital *humanists* we know that a crucially important dimension of that representation is precisely the disciplinary norms we adopt, and these we know to be founded on debate rather than on straightforward agreement. We can assume, in other words, that our future collaborators will disagree with us on some fundamental point, while nonetheless wishing to make use of what we have done. So the question then is: what needs to be expressed in order for this kind of longitudinal collaboration—one in which the first collaborator has to play their full hand before the other even comes into the game—to work? Is there any conduit through which such a negotiation can take place?

There is, but it takes a complex form. What is needed in this case is not simply a common language for data representation, but more importantly a common mechanism for *negotiating about data*

---

[2] See http://digital.lib.ucdavis.edu/projects/bwrp/.

[3] NINES is the Networked Infrastructure for Nineteenth-Century Electronic Scholarship, http://www.nines.org.

representation. And that, I would like to argue, is exactly what the TEI properly provides: not a perfect language for collaboration, but a—let us call it a functional—mechanism for negotiating about language: in fact, a mechanism for negotiating dissent.

TEI Customization

In order to give some concreteness to this assertion, we need to take a brief excursus into the details of the TEI customization mechanism, which is the form given to this negotiation process. The TEI encoding language is not a single, unitary thing: it is a vast landscape of possibility that covers domains as remote as manuscript description, dictionaries, drama, oral history, linguistic corpora. The TEI community as a whole needs to be able to represent all of these things, but any individual project needs just a selection: manuscripts and drama, say, or verse and scholarly editing. So the TEI language is represented at the most basic level as a set of possibilities: a single source file (called the "ODD"[4]) that represents the entire landscape of the TEI in potential terms.

To represent our own, more selective view of that landscape, we need another piece of information, a separate file which is called an ODD customization. The ODD customization file represents the world of an individual user or project: the set of choices through which the individual adapts the TEI schema to local usage. These choices may simply be selections from among the various parts of the TEI (elements to encode manuscripts, elements for scholarly editing, elements for representing dictionaries or drama or verse). They may also take the form of added constraint: for instance, a project may wish to limit the vocabulary used to describe poetic genres. But they may also be extensions: added elements for things that the TEI cannot yet represent, or changes to TEI elements to reflect local needs.

From these two files (the main TEI ODD, and the ODD customization), with appropriate processing, one can then generate a schema that expresses the TEI landscape as viewed through the lens of the individual application. And by comparing any two (or more) customization files, one can also in principle gain an understanding of how two projects differ in their representation of data. There are thus several potential vectors of difference or dissent that can be expressed here: between the individual researcher or project and the TEI community as a whole (expressed through actual changes and extensions to the unmodified TEI language), and also between different individual researchers or projects.

To give a concrete example, let us take the Women Writers Project again: a long-standing TEI project with a somewhat idiosyncratic set of texts, experimental views on text encoding, and a strong collaborative instinct. The WWP's TEI customization captures a number of important things:

- the specific sections of the TEI that the project actually uses (for instance, the TEI module for names and dates, the module for linking)
- the controlled vocabularies used in classifying things like verse structures

[4] The origin of this terminology is whimsical: because the ODD file represents both the schema itself and the documentation of the schema, "one document does-it-all".

- the changes the project has made to individual textual structures, to accommodate the idiosyncrasies of the WWP's texts or of their thinking about them

So in this case, the customization immediately serves two important functions: first, to express and document the rationale behind the WWP's encoding practice; and second, to express in formal terms the relationship between that practice and the TEI guidelines themselves.

Imagine, though, that in a few decades, after the project has run its course and its data is part of the Brown University repository, another group of researchers (perhaps working on women's writing in French) wants to use this data, in combination with their own, to compare ideas of genre across languages. In the wild, this project is utterly hopeless. But if both groups are using the TEI, the process is eased: the peculiarities of each project's encoding are expressed as points of dissent from the TEI itself: in other words, from a known quantity. In areas where the WWP does not disagree with the TEI, it uses the standard terms and structural concepts. The second project, coming later, is able to use the standard TEI markup present in the existing project data unmodified; from the WWP's ODD file they can also see explicitly where this encoding dissented from the TEI consensus, and assess how these points of difference accord with their own methods (and modify the WWP data if necessary). They can also bring in data from other TEI projects on the same basis: because all of the dissent has been organized into a set of vectors all pointing in towards the same hub (instead of being expressed combinatorially), the challenge of harmonizing the data and discerning the points of essential agreement and disagreement is significantly diminished (though of course not eliminated).

What is taking place here is a form of negotiation that manages in a subtle and bizarre way to be bi-directional even though it takes place across the passage of time. Each project, through its TEI customization, places its data in an explicit relation to the standard, with any reservations or qualifications clearly expressed: thereby deliberately putting the data "up for collaboration". What is distinctive about the TEI customization mechanism is that it formalizes dissent in this way, in relation to the standard, and allows its vector to be traversed in two directions. The same path that leads away from the unmodified TEI standard towards the individual application (from generality to specificity) can also be followed back to the center again. This traversal can be effected both by human beings and by computer processes. Information concerning what has been changed and why can be expressed in human-readable form and may serve as a valuable support in understanding the methodological choices that underlie a project's encoding practice.

Similarly, the ODD customization file can serve as the basis for automated analysis of difference and similarity of encoding methods. By analyzing a set of customization files, for example, we could identify all projects from a large set that use the same set of TEI modules or remove the same set of elements. We could generate a list representing the greatest common set of values for a given vocabulary (such as poetic genres) across a group of projects and also identify the values that are unique to each project, or identify the range of new elements created by each project, together with their closest TEI equivalents.

Interoperability and Collaboration

Taken as a whole, the customizable approach taken by the TEI permits the standard to function (both socially and technically) as an agreement at many levels—on the intention to treat data as a sharable and preservable resource, on the value of shared data standards, on the descriptive utility of this particular approach to modeling humanities texts, and on the impossibility of creating a single descriptive model that will satisfy all needs.

This last point may seem like a significant or even fatal concession, and it may be appropriate to consider here whether collaboration, in the absence of a single descriptive model, is really possible at all. Would it not enhance the collaborative effect in the examples given earlier, for instance, if instead of the elaborate articulations of the customization mechanism we had a simple and uniform standard for exchanging data in an unambiguous manner? Is this not in fact the mission with which the TEI was initially begun? Martin Mueller put the case for the value of uniformity in a presentation at the 2008 TEI conference:

> To compare is to look for significant difference, but without comparability difference cannot be identified in the first place….consistent encodings across many texts in a heterogeneous document space have a greater scholarly pay-off than finely grained encodings in closely defined but not necessarily compatible environments. (Mueller 2008)

From this perspective, in which the goal of interoperability is given primacy, the heterogeneity of different encoding approaches looks simply like grit in the works.

The challenge here is to understand the relationship—and the difference—between interoperation and collaboration. Interoperation is a functional property of data and tools: a measure of how seamlessly a new data set will perform within a system not explicitly designed for it. If I have written a stylesheet that expects a certain kind of TEI data, and your data flows through it without producing error messages or an unprocessable residuum, then your data is interoperable with my stylesheet.

Or is it? What if our measure of seamlessness here is not the unimpeded working of the tool, but the meaningfulness of the output? Take as an example a stylesheet that is designed to generate an index of first lines in poetry. It looks for the first "verse line" child of each "poem", extracts it, and sorts the resulting list. Imagine now that my collection of poems includes some with epigraphs or introductory notes following the poem title. If the schema for poetry does not include separate markup elements for these features and I cannot alter it, I may be compelled to represent them as part of the poem itself, using the element designated for verse lines. The stylesheet, none the wiser, obediently identifies these and extracts them as "first lines" without complaint, and the resulting index contains a mix of first lines and things that are not first lines but cannot be distinguished from them. From the standpoint of interoperability the error is hard to detect. The data itself is structurally indistinguishable from what the system anticipates; the problem is simply that the data is lying.

This problem is not a belated discovery, but a deep tension that has inhabited the design of the TEI from the start. Let us now recall, from earlier on, Poughkeepsie Principle 1: "The guidelines are intended to provide a standard format for data interchange in humanities research." This seems

like an unambiguous statement about the primacy of interchange and its dependency on concepts of standardization. But in the "Design Principles for Text Encoding Guidelines" which comment and expand on these principles, the editors of the TEI Guidelines note concerning this point:

> For interchange, it must be possible to translate from any existing scheme for text encoding into the TEI scheme without loss of information. All distinctions present in the original encoding must be preserved. (TEI 1988)

Meaningful interchange and interoperability thus require us to attend not only to the structural uniformity of our data, but to its semantic uniformity as well[5]: in other words, the distinctions being preserved in one encoding must be mapped onto the distinctions that are relevant in another encoding. The brief admonition that "all distinctions present in the original encoding must be preserved" is in fact a statement in support of the distinctiveness of that "original encoding" and the legitimacy of its distinctions: the role of the interchange format is to permit translation, not to efface difference. As Sperberg-McQueen's admonition to Olsen insists, uniformity comes only at the cost of eliminating what is "absolutely essential for dealing rationally with some texts." Fifteen years after that exchange, with digital editing now an important application of the TEI, we might expand this point to read "…and for dealing articulately with some editorial problems." The uniformity that cannot be assumed in our primary sources is equally rare in our scholarly methods. To collaborate effectively under these circumstances is thus a matter not of enforcing an artificial uniformity through which vital distinctions are elided, but rather of supporting the real and accurate exchange of the data in which we have a strong stake. The fundamental question, in other words, is whether we really want to hear what our collaborators really have to say.

This is not an argument for diversity for its own sake. In a strong collaborative ecology, the corollary of respecting each others' differences is respecting the commonalities that draw us into joint work in the first place. Structurally speaking, the TEI customization mechanism is as well suited to creating schemas that reflect a well-honed, tightly constrained expression of those commonalities as it is to expressing individual divergences from the consensus. But while a well-constructed schema can represent and support a consensus of this kind, it cannot create it. A standard like the TEI, in other words, is not the appropriate mechanism for *producing* a successful collaborative agreement except in cases where the quality of the data (its semantic nuance or its methodological rigor) is of no consequence. It is, rather, an extraordinarily effective mechanism for *representing* such an agreement once it has been made by the collaborating parties.

This point complicates what is commonly assumed about data standards in the digital humanities: that they support collaboration unproblematically, and that the better we standardize, the stronger our collaborations will be. Representational systems like the TEI operate in a domain where our descriptions, in their complexity and their consequentiality, reach towards the condition of human language and thought. This complexity is a strength and a measure of the intellectual potential of these representations, not a weakness. The challenge—as with human language—is to achieve mutual intelligibility while still being able to say what is worth saying. Collaboration, too, walks

---

[5] I am indebted to my colleague Syd Bauman for sustained discussions of this point that have helped me understand its significance.

this precarious line between egoism and altruism, between private insight and public communication, between local nuance and common ground. In this respect, the TEI is not simply a tool for collaboration but a methodological reflection of its structure at the deepest level.

Works Cited

Allen, C. 2007. Guidelines for Supporting ISO Code Sets. Available at: http://ns.hr-xml.org/2_5/HR-XML-2_5/CPO/GuidelinesForISOUtilities.html [accessed: 20 May 2010].

British Standards Institute. 2010. What are the differences between consensus and commissioned standards? Available at: http://www.bsigroup.com/en/Standards-and-Publications/About-standards/Differences-between-Consensus-and-Commissioned-standards/ [accessed: 20 May 2010].

Burnard, L., Bauman, S., eds. 2007. *Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium. Available at: http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index-toc.html [accessed: 20 May 2010].

ISO. 2004. Information technology — Codes for the representation of human sexes. Available at: http://standards.iso.org/ittf/PubliclyAvailableStandards/c036266_ISO_IEC_5218_2004(E_F).zip. See also http://en.wikipedia.org/wiki/ISO_5218 [accessed: 20 May 2010].

Mueller, M. 2008. TEI-Analytics and the MONK project. TEI annual conference, London, November 6–8, 2008. Available at: http://www.cch.kcl.ac.uk/cocoon/tei2008/programme/abstracts/abstract-169.html [accessed: 20 May 2010].

Nunberg, G. 2009. Google's book search: a disaster for scholars. *The Chronicle of Higher Education.* [Online] August 31. Available at: http://chronicle.com/article/Googles-Book-Search-A/48245/ [accessed: 20 May 2010].

Olsen, M. 1997. Humanist Discussion Group, Vol. 10, No. 895. Available at: http://www.digitalhumanities.org/humanist/Archives/Virginia/v10/0515.html [accessed: 20 May 2010].

Sperberg-McQueen, C. M. 1995. Preface to TEI Bare. [Online] Available at: http://www.tei-c.org/Vault/Bare/ [accessed: 20 May 2010].

TEI. 1988. Design principles for text encoding guidelines." TEI ED P1. Available at: http://cmsmcq.com/1990/edp1.html [accessed: 20 May 2010].

Trimbur, J. 1989. Consensus and difference in collaborative learning. *College English* 51.6 (October 1989), 602-616.