Evaluating the *Bergen Electronic Edition* | Herbert Hrachovec

1. The Bergen edition and Wittgenstein scholarship

Current Wittgenstein scholarship is marked by a striking anomaly. The *Bergen Electronic Edition*, which was published in instalments beginning in 1998, has dramatically changed the field of Wittgenstein philology. Wittgenstein's entire writings are now available in easily accessible facsimiles as well as in carefully prepared diplomatic and normalized transcriptions. This is nothing less than a quantum leap for anyone involved in going beyond the surface of those volumes of the *Nachlass* published by the Trustees, some of which have been shown to be in need of philological revision. The search facilities included in the Bergen edition are unique in providing almost instant access to all the data parsed by arbitrary queries. The very scope of the enterprise, offering a comprehensive, multi-layered digital rendition of the Wittgenstein corpus goes far beyond anything we can expect in our lifetime from traditional editions, including Michael Nedo's *Wiener Ausgabe.* And yet – and this is the anomaly – a significant number of recent books on Wittgenstein do not even mention the Bergen edition.

The New Wittgenstein, a collection of essays published in 2000[1] contains a bibliography that faithfully reproduces all *primary sources*, but lacks any reference to the digitised *Nachlass*. *Wittgenstein in America*, a prestigious collection from 2001 – ironically published by Oxford University Press[2] – does no

---

1. Alice Crary and Rupert Read (eds.): *The New Wittgenstein*. London and New York. Routledge 2000

better, and the same holds for the German language literature. To pick just two examples: neither Eike von Savigny's reader on the *Philosophische Untersuchungen*,[3] nor Wilhelm Vossenkuhl's corresponding volume on the *Tractatus*,[4] contains any pointer to the Bergen project. Clearly something strange is going on here.

The easy explanation is that decades of Wittgenstein scholarship have grown accustomed to the printed sources. It is just a matter of time before philosophers become aware of the additional resource. Another likely reason is that most of the newly available material is in German and will, therefore, not immediately appeal to the overwhelmingly anglophone Wittgenstein community. Both explanations certainly sound plausible. But the focus of this paper will be on some shortcomings of the digital edition that may partly explain why the innovative work done at the Wittgenstein Archives has received so little recognition. In order to discuss this topic I shall have to deal with issues that go beyond the scope of Wittgenstein philology proper. As it turns out, the Bergen project raises some fairly general questions about the socio-economics of computer-assisted scholarship. It is only when we consider some of the conditions that current digital technology imposes upon the humanities that we are able to notice – and hopefully correct – a certain weakness of the Bergen approach. The first part of this paper attempts to give an outline of the overall problem, whereas the second will present ongoing research to address some desiderata revealed by the preceding analysis.

2. Technical and other troubles

The Vienna story

Here comes the story – sad, but true – of how the Bergen edition vanished from the data-bases available to members of the University of Vienna. In

2. Timothy McCarthy and Sean C. Stidd (eds.): *Wittgenstein in America*. Oxford. Clarendon Press 2001

3. Eike von Savigny (ed.): *Ludwig Wittgenstein: Philosophische Untersuchungen*. Berlin. Akademie Verlag 1998

4. Wilhelm Vossenkuhl (ed.): *Ludwig Wittgenstein: Tractatus logico-philosophicus*. Berlin. Akademie Verlag 2001

1999 the University library acquired a network license from Oxford University Press and made texts and facsimiles available via its campus CD-ROM server. Based on an MS-Windows NT system, the server actually used software supplied by Citrix, a company offering free service under various operating systems, to match their CD-ROM host software. Consequently, MS-Windows-, Mac- and Unix-based users could access the Wittgenstein *InfoBase*. In the summer of 2001, however, the university's CD-ROM server was transferred to a different environment, which sadly resulted in loss of access to the Bergen edition. Two months of gentle prodding did not help a bit, so I decided to investigate the matter. The initial move, which had a number of consequences, involved an organizational switch. Responsibility for maintaining the university's digital archives passed from the library to the computer service centre because of the increasing complexity of installing and maintaining a great number of database applications on a campus network. An interview with the IT professional in charge introduced me to a veritable clash of traditions. As the engineer put it: "The library people want to have some booklet or box onto which to put a label." His own preference was completely different. Rather than worry about how to smooth out the incompatibilities between conflicting software drivers for a considerable number of applications, updated at different intervals, his preferred option was simply to plug in at the site of the original data provider, who is presumably most competent in handling the information. This procedure would spell the end of burning and mailing physical CDs, in other words, of treating them as analogous to books rather than as information deposits.

This predisposition led to a slack attitude when it turned out that the Oxford CDs could not easily be installed in the new environment. In fact they kept crashing after a few minutes, prompting the engineers to suspect a software bug or, alternatively, defective CDs. The difficulty is as yet unresolved. My aim is not to voice a general warning against the pitfalls of information technology. At issue here is something far more specific, which bodes ill for humanities scholarship. Excuse me if I have to delve even deeper into seemingly anecdotal details. Like it or not, such details are of enormous importance in facing the challenge of future electronic philosophy. I shall mention and briefly discuss three areas of conflict highlighted by

the experience at Vienna University. In abstract terms these are: scarcity of resources, market economy and the dynamics of software development.

Scarcity of resources

Books do not demand a lot of attention once acquired and put on a library shelf. It has become clear that this is not the case with digital data dependent on the employment of computers. Scholars find themselves trespassing on unfamiliar territory. The speed and scope of networked information sharing are certainly convenient, but some disturbing developments have already been hinted at. Only so many applications run smoothly on many existing CD-ROM servers. Whereas in the old days a library had only to provide storage facilities, digital philosophy finds itself in competition with vastly more popular resources, backed by more powerful interest groups, which are, in turn, equipped with substantial funds to pursue their aims. It does not need much imagination to figure out the loser if a conflict between a database serving the Department of Medicine and the Bergen edition should ever arise. When we consider that in fact all general interest databases are considerably more important to university administrations and that they are regularly enlarged, involving possible software conflicts at every update, the sudden disappearance of a relatively minor textual resource is not at all surprising. And before you complain you should remember that the person responsible might well confront you with last year's statistics showing exactly how many colleagues used the electronic law library as well as the citation index. Sceptics used to argue that the administration of the Vienna Festival could save a lot of money if it simply bussed the complete audience of a particular program to whatever city the play invited to the festival was originally produced in. It is not unlikely that providing humanity scholars with their personal copy of the electronic corpus would prove a cheaper alternative to sharing a common information network.

Market economy

There is a second source of pressure on the idea of fair and equal distribution of electronic knowledge among the community of investigators. I have mentioned the Citrix server originally employed by the university library. Now, as it happens, Microsoft has more or less taken control of Citrix, repositioning the product. Advanced Microsoft operating systems are to include

a CD-server of their own, while the Citrix software is destined to cover the high end of the market. The base line is that an all–Microsoft client-server environment is made considerably more attractive, whereas people using any other operating system have to pay extra for the add-on Citrix solution. Corporations are expected to make profits, so one should not be too surprised about such moves. Yet, they are somewhat disquieting from the point of view of traditional scholarship, which is shown to be at the mercy of market forces that control the very prerequisites of its labour. The marginalization of minority interests is a clear case of capitalist economics spilling over into the academic world. The prospect that the future course of digital philology will be determined in Redmond should make everyone involved more than uneasy.

Dynamics of software development

But, even accepting this situation, one more problem is revealed by the Vienna episode. The Bergen edition depends not only on a MS–Windows environment. Its entire content is put into a software envelope called *Folio Views*, which is intended to ease use of this considerable amount of data. Folio Corporation is a commercial enterprise too, or, to be more precise, it *was* a commercial enterprise until it was taken over by NextPage. This is how Folio customers are wooed on the NextPage site:

"You've relied on Folio technology for years. It's taken you where you need to be today. But what about tomorrow? As e-business moves to the Web, how will you fare against your competition?"[5]

This does not have the flavour of academic pursuits, to put it mildly. And NextPage would not be helpful anyway, since the Folio Views version used by the Bergen edition is 3.11, whereas NextPage has just discontinued support for versions 4.21 to 4.23. The functionality of the Wittgenstein CDs is obviously not affected by such developments in the business of archiving software. Still, this is a matter of concern for the future. If the Bergen edition is to be adapted there will be no Folio Views to meet the demand of current technology. A new decision will have to be taken and it has become

---

5.   http://www.nextpage.com/folio. Accessed November 15, 2001

obvious how deeply such decisions are affected – and in turn affect – some basic presuppositions governing social control of information technology.

Software restrictions

Since my aim is to explain the reluctance of Wittgenstein scholars to embrace the Bergen edition, it may well be objected that the discussion so far has dealt with details that can hardly account for this attitude, if only because trouble with the CD server at Vienna university is much too local an incident. True enough, yet my contention is that a vague awareness of this type of difficulty leads people to shy away from actually involving themselves with the digital Wittgenstein *Nachlass*. As my account has shown, such apprehensions are not entirely unfounded. One needs a robust faith in technology in the face of some obvious deficiencies to opt for an electronic Wittgenstein. I shall conclude this section by elaborating on some of the constraints that Folio Views imposes upon scholars. The format prescribed by this particular software package is, it seems to me, another reason for scepticism among our academic colleagues.

The MS-Windows rendition of Wittgenstein's writings has been encoded into large binary files measuring tens of megabites in size. The only access to textual data is via the graphical user interface provided by Folio Views. Several reasons for this arrangement can be given. Putting the files into binary code adds speed and makes for very easy searching. It also protects the data from unlicensed manipulation since one has to buy the whole package to get at any particular Wittgenstein text. You can extract the information and save it in so-called *shadow files* which allow you to copy and paste text and many other functions. For most purposes of standard exegesis the Bergen edition is an excellent tool, providing a complete set of facsimiles, two carefully edited versions of the underlying material, superb search facilities and tracebacks, as well as a copy and paste mechanism. This is considerably more than you can expect from any printed source. To notice some of the shortcomings one has, in fact, to consider *digital* alternatives to the present format.

Books and printed documents can be physically arranged at will. This freedom is usually echoed by icons that can be moved around the virtual desktop. Folio Views does not offer this kind of mobility but rather joins one manuscript after the other into one single compendium with only a table of contents to direct users to particular volumes. This is an awkward

way to start working on selected sources, yet it is the only one available if you lack the permission to create and modify shadow files (which can be the case if you work over a network). Arranging the items in numerical order according to the von Wright standard raises a further problem, since the numerical sequence of the *Nachlass* volumes does not coincide with their chronological genesis. Typescript TS 201a from 1913 is preceded by notebook MS 140 from 1934 just because of the von Wright numbering. This is irritating for searches since the result usually lacks chronological consistency. While it would certainly be too much to expect the editors to deal with the delicate question of temporal interdependence of Wittgenstein's manuscripts, it seems fair to demand the freedom to put those virtual volumes into any order one finds appropriate for a given purpose. This is made unreasonably difficult by putting them into the straightjacket of Folio Views.

One final observation prepares the ground for the second, more constructive, part of this paper. As far as I could determine, extracting text from the Folio Views InfoBase has to be handled with care. Features like italics, underscores etc. can get lost, whereas hidden code, i.e. dates and page numbers, are by default inserted into the ASCII output. One has to re-normalize every extract. There is a perfectly good reason for the loss of information: ASCII is the lowest common denominator across existing computer platforms and it simply does not yield the finer distinctions needed by more advanced typesetting. Yet the situation described is somewhat paradoxical. Since users are forced to use MS-Windows and Folio Views to access Wittgenstein's text anyway – why not offer a format that preserves the original information and is suitable for a MS word processor? There is a misfit between the two InfoBases offering one preset view each and the material put at the reader's free disposal. In general, quoting Wittgenstein from the Bergen edition by copying and pasting his text when accessing it over a network, can be difficult. This seems a very unsatisfactory situation for such an expensive product. It has to be admitted, though, that there is more to this issue. The problem indicates a more general difficulty and calls for a second look at the Bergen project, taking into account the background of electronic philology barely mentioned so far.

3. Prospects with XML

Criticism of digitised text faces a dilemma. If such text were required to achieve general cross-platform compatibility on all available computer systems it would be forced to use ASCII code. But this is unacceptable, since this code lacks even the most basic typographical conventions needed by a philologist. A simple concept like *quotation*, to pursue the previous example, is transformed into a software construct on a WYSIWYG virtual page. While ordinary scholarly quotation is insensitive to the peculiarities of paper, ink and print, this is no longer the case where computer generated *pages* are concerned. The first part of this paper is in fact an elaboration on this crucial point. Computer systems, convenient as they may be for scholarly purposes, introduce entirely new and partially disturbing factors into the field of philology. One of the greatest challenges is to resolve this dilemma, and it is here I find the CD edition a somewhat unconvincing compromise between the requirements of highly professional criticism and highly volatile media tools. Is there a better way to approach the inherent conflict between long-term standards of independent scholarship and the market pressures that affect the required software equipment? The answer is a resounding *yes* and, furthermore, it is a cue to take a closer look at what the Bergen Wittgenstein project has actually achieved.

The digital *Nachlass*, as edited in Bergen, resists the scepticism just expressed, although one would hardly think so by looking at the monitor. Electronic scholarship has found a solution to the dilemma described above. To put it very simply: use ASCII meta-code to indicate the desired additional information within straight ASCII text. A so-called mark-up language does not try to render italics on the screen of the end user. There *is* no single way to achieve this, given the plurality of digital interfaces. Rather than attempting to please a transient majority of readers a scholarly mark-up language captures philological content in meta-tags and does not involve itself in questions of presentation. The down side is that this does not give you – for example – italicised text on *any* platform. It simply indicates that a certain sequence of characters should be italicised, or put into a footnote, or omitted from the final version. This caution is, on the other hand, a crucial move to win independence from the software requirements of the day. A two-step procedure, as envisaged by mark-up languages, defers the satisfaction of immediately dealing with virtual mirror pages of any given page. But

it preserves the autonomy of scholarship against the flux of digital consumer economy. And it is this approach that guides work at the Wittgenstein Archives. The Folio Views product is just one instance of a vastly more extensive *corpus* of information coded into the so-called source transcriptions. It is here that things begin to get interesting.

On the one hand we have transcriptions of the textual evidence into a sophisticated mark-up language (MECS), which preserves every step of Wittgenstein's work flow by means of complex constructions in a technical language. At the other end of the spectrum users are given two fairly rigid views of the Wittgenstein *Nachlass*, building upon programs that are presumed to be user friendly at a given time. There *has* to be a software bridge between marked-up code and something philosophers can actually read on their machines. But it is by no means necessary to use Folio Views, or any other commercial product that is bound to undergo alterations due to forces beyond the reach of academia. Instant 1:1 correspondence between facsimiles and this year's technology is, in fact, the wrong way to go. It is, of course, a time-honoured and very gratifying state of affairs in the world of printed books, witness the splendid edition of the *Philosophical Investigations* by Joachim Schulte et al. published in 2001.[6] Yet computer texts should not attempt simply to mimic printed originals. Electronic philology loosens the grip traditional books hold upon our imagination. It is crucial to notice that the new presentational medium offers considerably more flexibility in conveying change within its subject matter and of changing the medium itself. A monitor is not a printed page and it is precisely because of the software bridge that mediates between source transcriptions and WYSIWYG output that the cluster of problems I have presented in the first part of this paper arises. Even though the Bergen edition has to satisfy the expectations of scholars reared on the Gutenberg Galaxy the project team would be ill advised to aim for just books in digital disguise. Attention has to be directed towards the software mechanism in order to reveal the full potential of computer-aided philology.

---

6. Ludwig Wittgenstein *Philosophische Untersuchungen. Kritisch-genetische Edition*. Herausgegeben von Joachim Schulte in Zusammenarbeit mit Heikki Nyman, Eike von Savigny und Georg Henrik von Wright. Frankfurt/M 2001. Suhrkamp

So, what are the alternatives to filtering the source transcriptions into the present mould? Since they are subject to a certain well-defined grammar they can, in principle, be translated into any desired additional format. One rendition is, however, of special importance to our present purpose. The Wittgenstein Archives and Claus Huitfeldt are working on a MECS-to-XML converter, the availability of which will have a decisive impact upon the present editorial arrangements. The reason is that such XML documents, unlike those we have at the moment, can be used by everyone, irrespective of designated operating systems and word processors. Such documents, it is true, do not provide an isomorphism to the underlying originals that you could recognize at a glance. Reading the source transcriptions is like listening to a theatrical performance in which all the stage instructions are verbalised. XML is itself a mark–up language, enabling its users to capture the relevant features by way of meta–data as described before. The crucial difference to MECS is that the XML standard is widely popular and that there are numerous commercial as well as open source applications that allow users to extract, rearrange and further process XML-encoded information.

Notice the difference between source code distilled into the format of some particular word processor and translated into XML. All the convenience of being able to work immediately on the text is lost in the second case. Yet this is the price one pays for a significant improvement in the general scholarly setting. With XML, dependence on the specifics of particular machines is minimized and one can choose one's own way of processing the data. I should immediately add that this is not something one would expect the average reader of Wittgenstein's *Nachlass* to do. There is an indubitable need for the CD edition in its present form. But the points made about its rigidity are not just theoretical complaints. They are mentioned in order to prepare the ground for a broader vision of digital transcriptions. Documents coded in XML provide platform independent patterns of textual information which can be enriched with suitable content and without loss of generality. To illustrate these challenging opportunities I turn finally to an international research project entitled "Tracing Wittgenstein: Digital Explorations" (http://wittgenstein.philo.at, accessed November 1, 2004).

"Tracing Wittgenstein" is working with XML (and HTML) versions of manuscript 115, which are publicly available from the Bergen archive. One editorial improvement that many of Wittgenstein's collations seem to call for

is some guidance to the overall structure of the assembled remarks. The need for some table of contents was felt, for instance, by Rush Rhees, whose 1964 edition of the *Philosophical Remarks* starts out with an extensive tableau briefly describing the contents of the manuscript in sequence. While this is certainly a helpful addition, Rush Rhees goes on to violate some basic rules of textual criticism by superimposing his own accounting system upon Wittgenstein's collection, mentioning only in passing that none of this is to be found in the original text. It seems obvious that a critical edition must refrain from such beautifications of the evidence, although most people will still want to be given a general idea of what the author is up to at any given point. Traditionally, introductory and exegetical writing has tried to provide such help. One fairly simple thing one can do, given an XML version of one of Wittgenstein's original sequences of remarks, is to adjoin them to a tree-like representation of some table of contents. This is already implemented in one of the outcomes of the project which can be down-loadad from http://wab.aksis.uib.no/wab_115ape.page (accessed November 1, 2004). The branches of this tree, in other words the sections, chapters and further subdivisions one's hermeneutics has produced, can serve as handles to access the underlying material which, at the same time, is preserved without inappropriate interference. This strategy seems to differ very little from well-known hermeneutical procedures. But make no mistake; it opens up some options hitherto unavailable within the academic world.

One comparatively moderate enhancement is the ability to regard one's involvement with Wittgenstein's text as an ongoing, public enterprise. One does not have to come up with more or less definitive results which are then put into print and preserved unalterably. Electronic structural analysis of the *Nachlass* is sensitive to peer criticism and can easily respond to suggestions and improvements from outside commentators. A second step suggests itself, and here we enter into a realm unprecedented in traditional book culture. Without much effort we can include *several* competing proposals for the proper account of the structure of the underlying remarks. This means that a group of scholars may cooperate, offering distinct views based upon the same textual material. Subdividing Wittgenstein's sequences into smaller units, designing different hierarchies and dependencies, is just a start, however. One or more commentaries can be run parallel to the text with any of them referring to further text, or commentaries, or additional outside infor-

mation by hyper-links. The *Nachlass* evidence will again remain outside such possible features, serving as the common point of reference for those digital add-ons. A more ambitious plan would be to extend the present mark-up to include semantic information. The development of Wittgenstein's discussion of *Zahnschmerzen*, to mention but one example, contains some remarks on *Magenkrämpfe*, which will be overlooked by anyone searching for the more prominent term. One or several scholars might develop a kind of *thick* description of (parts of) the *Nachlass* preparing the ground for more specific, individual philosophical work.

4. The Bergen edition and digital scholarship

In this paper I have not yet raised any question about Wittgensteinian philosophy and very much regret being unable to do so in conclusion, particularly since only a more detailed account of the minutiae of Wittgenstein's elaborations could convince a sceptical listener of the fruitfulness of the envisaged kind of exegesis. Suffice it to say that Wittgenstein's textual strategy turns out to be extremely subtle in his manuscripts. He is careful to arrange his remarks in such a way as to achieve *Übersichtlichkeit*, putting considerable weight on the structural arrangements of paragraphs to make his point. Wittgenstein's writing exhibits a musical quality, using repetition, inversion, contraposition and variation of thematic threads to explore the scope of his ideas on any given subject matter. It has long been recognized by commentators that the development of such conceptual patterns is a crucial feature of the philosophical activity as conceived by the author. We begin to become aware of the extensive array of cross-references and re-arrangements characteristic of the *Nachlass*. None of this can easily be captured in a once-and-forever edition. Conventional scholarship is called for to pick out the relevant leads and follow the traces of Wittgenstein's philosophical development. It has been done and obviously will continue to be done in print. I hope to have convinced the reader that a collaborative approach focusing on the as yet untapped potential of the source transcriptions is a new and worthwhile direction of research.

Books are two things in one: authors decide upon their content while editors put such contents into one particular form. The fluidity of thought in Wittgenstein's *Nachlass* does not fit well into hardcover bindings and the situation is not much better with respect to silver disks. Software developers

talk of a *feature freeze* to indicate that one has to set a limit at a given time and place to what can reasonably be achieved. This is how books get written and published, including *Nachlass* editions. It is probably not the best way to approach the on-going activity of philosophical argument and peer research. The challenge facing the profession is to come up with cognitive and institutional models that will further the use of digital technologies in enhancing that profession's long-term aims. A big step has been taken by putting together the Bergen edition. More steps remain desirable, releasing the dynamics inherent in scholarly digitisation.