

Evaluating the *Bergen Electronic Edition*

Herbert Hrachovec

Current Wittgenstein scholarship is marked by a striking discrepancy. The Bergen electronic edition, which has been published starting in 1998, is now completed and has dramatically changed the field of Wittgenstein philology. Wittgenstein's entire writings are 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 the volumes published from the *Nachlass* by the Trustees, some of which have been shown to require 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 from traditional editions, including Michael Nedo's "Wiener Ausgabe", in our lifetime. And yet – this is the discrepancy alluded to – a significant number of recent books on Wittgenstein does not even mention the Bergen edition. "The New Wittgenstein", a collection of essays published in 2000¹ contains a bibliography faithfully reproducing all "primary sources", but lacking any reference to the digitized *Nachlass*. "Wittgenstein in America", a prestigious collection from 2001 – ironically published by Oxford University Press² – does not do better and the same situation 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 contain any pointer to the Bergen project. Something strange is, clearly, going on here.

An easy explanation offers itself: decades of Wittgenstein scholarship have been accustomed to the printed sources. It's just a matter of time until philosophers become aware of the additional resource. Another likely reason is the fact that most of the newly available material is written in German and will, therefore, not immediately appeal to the overwhelmingly Anglophobe Wittgenstein community. Both explanations certainly sound plausible. But, talking in Bergen, I will not waste my time in preaching to the converted. The focus of this presentation will rather be on some shortcomings of the digital edition that may be partially responsible for the amazing lack of recognition the innovative work done at the Wittgenstein archives has received. In order to discuss this topic I'll have to deal with some issues outside the scope of Wittgenstein philology proper. As it turns out, the Bergen project raises some fairly general questions pertaining to the socio-economics of computer-assisted scholarship. It's only against the background of several conditions imposed upon the humanities by the current implementations of digital technology that a certain weakness of the Bergen approach can be apprehended and – hopefully – corrected. The first part of this talk attempts to give an outline of the overall problem, whereas the second one will present ongoing research to address some desiderata revealed by the preceding analysis

1 Troubles

Here comes the story – sad, but true – of how the Bergen edition vanished from the data-bases on offer for members of the University of Vienna. Starting in 1999 the University library had 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 clients running under various operating systems, to match their CD-ROM host software. MS-Windows-,

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

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

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

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

Mac- and Unix-based users could, consequently, access the Wittgenstein *InfoBase*. This summer, however, the university's CD-ROM server was re-installed in a different environment and – alas – the Bergen edition had dropped out. Two months of gentle prodding did not help a bit, so I decided to investigate the matter. The initial move, triggering a series of consequences, had been an organizational switch. Responsibility for maintaining the university's digital archives had passed from the library to the computer service center 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 to simply 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: treating them in analogy to books rather than regarding them as information deposits.

This predisposition led to a certain amount of slack 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, but my aim is not to voice a general warning against the pitfalls of information technology. A much more specific arrangement, boding ill for humanities scholarship, is at issue here. Excuse me if I have to delve even deeper into seemingly anecdotic 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. To put it in abstract terms: scarcity of resources, market economy and the dynamics of software development.

1.1 Scarcity of Resources

Books do not demand a lot of attention once they are 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 in unfamiliar neighborhoods. Speed and scope of networked information sharing are certainly convenient, but some disturbing developments have already been hinted at. There are only so many applications running 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 that 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. Taking into account 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 remember, before complaining, that the person responsible might well confront you with last year's statistics showing exactly how many colleagues have been using the electronic law library as well as the citation index. Skeptics used to argue that the Vienna festival's administration 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's not unlikely that providing humanity scholars with their personal copy of the electronic corpus will prove to be a cheaper alternative to sharing a common information network.

1.2 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 over Citrix, repositioning the product. Advanced Microsoft operating systems are to include a CD-server of their own while the Citrix software is going 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 controlling the very prerequisites of its labor. The marginalization of minority interests is a clear case of capitalistic economics spilling over into the academic world. The fact that the future course of digital philology is going to be determined in Redmond should make everyone involved more than uneasy.

1.3 Dynamics of Software Development

But, even accepting this situation, one more problem is revealed by the Vienna episode. The Bergen edition does not only depend on a MS-Windows environment. Its entire content is put into a software envelope called "Folio Views", which makes for the ease of 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 being 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

It's not the flavor of academic pursuits, to put it mildly. And NextPage would not be helpful anyway since the Folio View 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 meeting the demand of current technology. A new decision will have to be taken and it has become obvious how deeply such decisions are affected – and in turn affect – some basic presuppositions governing social control of information technology.

1.4 Software Restrictions

I have set out to explain the reluctance of Wittgenstein scholars to embrace the Bergen edition and you might very well object that the discussion so far has dealt with details that can hardly be the reason for this attitude, if only because some trouble at the CD server at Vienna university is much too local an incident. True enough, yet my suggestion is that there is a vague awareness of this type of difficulties that 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 belief in technology in the face of some obvious deficiencies to opt for an electronic Wittgenstein. I will conclude this section with a number of remarks elaborating the constraints imposed upon scholars by Folio Views. 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 MBites 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. If you own a set of CDs you can extract the information and save it in so-called "shadow files" (This option seems, unfortunately, not to work over a network.) For most purposes of standard exegesis the Bergen edition is an excellent tool, providing a

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

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 and its the only one available if you lack the permissions to create and modify shadow files. 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 201a from 1913 is preceded by Notebook 140 from 1934 just because of the von Wright numbering. This is irritating for searches since the result will usually lack 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 talk. As far as I could determine extracting text from the InfoBase has to be handled with care. Features like italics, underscores etc. are 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 preserving the original information and suitable for a MS word processor? There is a miss-fit between two info-bases offering one preset view each and the material put at the reader's free disposal. One cannot, usually, quote Wittgenstein from the Bergen edition by copying his text. 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 up to now.

2 Prospects

Digitized textual criticism has to face a dilemma. If its results 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 pick up the previous example, is transformed into a software construct on a WYSIWYG virtual page. While ordinary scholarly quotation is insensitive against the peculiarities of paper, ink and print, this is no longer the case where computer-generated "pages" are concerned. The first part of this talk has, in fact, been an elaboration of 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 most prominent tasks at hand is a resolution of the dilemma indicated and it is here that I find the CD edition to be a not entirely convincing compromise between the requirements of highly professional criticism and highly volatile mediatic tools. Is there a better way to approach the inherent conflict between long-term standards of independent scholarship and market pressures operative upon the required software equipment? The answer is a resounding "yes" and, furthermore, it is a cue to take a closer look at what the Wittgenstein project has actually achieved.

The digital *Nachlass*, as edited in Bergen, escapes the expounded scepticism although you would not notice

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 – italicized text on *any* platform. It simply indicates that a certain sequence of characters should be italicized, or put into a footnote, or be omitted from the final version. This abstinence 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 actual 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 archive. The Folio Views product is just one instantiation of a vastly more extensive *corpus* of information coded into the so-called source transcriptions. Things begin to get interesting here.

We have, at one side, transcriptions of the textual evidence into a sophisticated mark-up language (MECS-WITT), preserving every step of Wittgenstein's work flow by means of complex constructions within a technical language. At the other end of the spectrum users are given two fairly rigid views onto 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 bound to be subject to alterations beyond the reach of academia. Instant 1:1 correspondence between facsimiles and this years technology is, in fact, the wrong way to go. It is, of course, a time-honored 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 this autumn.⁶ Yet, computer texts should not attempt to simply 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 mediating between source transcriptions and WYSIWYG output that the cluster of problems I have presented in the first part of this talk arises. Even though the Bergen edition has to satisfy the expectations of scholars reared on the Gutenberg galaxis 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.

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. Claus Huitfeldt is 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 verbalizing all the stage instructions. 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 enabling 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 immediately work on the text is lost in the second

⁶Ludwig 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

case. This is the price to pay for a significant improvement in the general scholarly setting, though. Given 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 nothing one would expect an 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. Their aim has been to prepare the ground for a broader scope of vision concerning 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 those challenging opportunities I, finally, turn to a research project proposed by Dieter Köhler and myself.

We are 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 as 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 in superimposing his own accounting system upon Wittgenstein's collection, just mentioning in passing that none of this is to be found in the original text. It seems obvious that a critical edition has to refrain from such beautifications of the evidence, but 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 a XML version of one of Wittgenstein's original sequences of remarks, is to adjoin them to a tree-like representation of some table of content. 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* proposals competing to give 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 information 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.

The prototype application designed by Dieter Köhler and philosophically attended to by myself, shows one way of implementing dynamic tables of content, based upon the close reading of an initial sequence of remarks from the diplomatic transcriptions of Ms 115. Here is a glimpse at what it looks like:

I have not yet, giving this talk, raised any problem in Wittgensteinian philosophy and very much regret to be 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 as to 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 place 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 an 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 you that a collaborative approach focusing on the 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 – at a given time and place – one has to put a limit at what can reasonably be achieved. This is how books get written and published, including *Nachlass* editions. It's 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 suitable to use digital technologies to enhance its 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 digitization.