

RESPONSE TO ROGER BAGNALL PAPER: INTEGRATING DIGITAL PAPYROLOGY*

Peter Robinson

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 3.0[†]

External Image

Please see:

http://rup.rice.edu/image/cnx_images/shapes-buybutton.jpg

A response might contain somewhere a list of areas in which the respondent disagrees, expresses reservations, or at least hints at dissent. I offer no such list. Bagnall's paper sets out superbly the state of play at a key moment in the movement of papyrological resources, from materials gathered by diverse scholars in many forms and expressed in print, through the increasing presence of digital methods and publication across a range of distinct projects, to an awareness that having one successful project is not enough; and then, a series of steps towards—well, we do not know what yet, but the word Community is writ large across the gate these projects are trying to pass through.

The paper does more than describe what is happening among papyrologists. Change some names, and a few references, and shift the date and the geography: the trajectory Bagnall sketches for papyrologists is exactly the same for at least three other groups of scholarly materials with which I am familiar. The Canterbury Tales Project, with its eighty-four fifteenth-century witnesses of the *Tales*; the *Commedia* Project, which has nearly finished work on seven manuscripts of Dante's *Commedia* and is contemplating with mingled fear and joy the other seven hundred ninety-three or so manuscripts; and then, the Greek New Testament projects in Munster and Birmingham, with some five thousand witnesses in Greek, and many more in many other languages. All contemplate the same landscape, with huge ranges of material suddenly accessible in digital form; with new models of collaboration and publication now available; with the same tensions between widening involvement and scholarly standards; and with the same asymmetry, of beautiful visions and scarce resources to achieve them. And, I am certain, it is not just the three of us, and the papyrologists. Many projects find themselves now, early in 2010, some twenty years or so into the digital access revolution sparked by the web, at the same point.

I will sketch out some more the points which make Bagnall's problems our problems. First, there is the volume of materials in many different media but now, increasingly, appearing in our browser: we all have manuscripts, papyri, catalogues, commentaries, dispersed across different media, times and places. Second, we have successful projects (at least, successful in that they did what they said they would do) going back a decade or more, which have created new masses of born-digital material. Third, we are not alone: a typical project both includes multiple partners within itself, and then partners with other projects. Fourth, we are aware that much as we might have done, we have barely started: in the Canterbury Tales Project we have published transcripts of only around 15 percent of the *Tales*, and all the rest of Chaucer, and then everything else in Middle English manuscripts lies before us. Fifth, we are all concerned about the future of

*Version 1.2: May 14, 2010 5:38 am -0500

[†]<http://creativecommons.org/licenses/by/3.0/>

what we have built so far. For many of us, it has been too personal a creation: who will put the same effort into continuing the work as we did into starting it? Sixth, we are discovering that traditional boundaries are dissolving in the digital world. Bagnall mentions the merging of text base and edition; the clear lines between transcription, editing, and reading are also blurring.

Seventh—and this is maybe the most crucial similarity, and where I found the Bagnall paper most pertinent—though our projects are led by scholars, and were originally conceived by scholars for scholars, we sense the presence of a wider context out there. We suspect that our work might reach thousands more, even millions. We suspect too that there are people in that audience who are more than interested, who more than wish us well: but who also have ability and knowledge to contribute to what we do—to transcribe, annotate, even edit. Every now and then something occurs to remind us of the potentials in this audience. On July 8 and 9 last year the British Library launched the Codex Sinaiticus website: in the first three days, over one million different people visited the site. Newspaper articles (mostly on the web, with CNN and Fox news leading the pack) about the project reached 200 million people worldwide. Very few of all those people could read a word of the manuscript: but there they all were.

This brings us to the one key word in Bagnall's paper I have already flagged above: community. Up to the last pages of his paper, Bagnall uses the term more narrowly than I have used it in the previous paragraph. Primarily, he is there thinking of a community of fellow experts: scholars and students in the academy interested in, and working on, papyrological materials who are not directly involved in the nexus of projects he describes. Let us dwell on the concept of a scholarly community, as invoked by Bagnall in the first part of his paper. Of course, what defines a community is what the people in it have in common. In a neighborhood, a town, or a city, a community is just people living in the same geographical area. We are familiar too with geographical communities containing many other communities, defined by shared church, or school, or age group, or interests, or political affiliation, and we know that these may overlap and extend across other geographical communities. Most recently, we have online communities, some of them vast.

Successful communities need more than shared interest: they need agreed rules. Over the centuries, academic communities have evolved their own rules, covering issues such as plagiarism, credit, publication and control. It would be pleasant if these same rules just carried straight over into the digital world, and indeed some do. But some old rules do not apply, or need drastic modification, and we need some new rules.

Bagnall touches on several of the rules. In what follows, I will elaborate these further, building on his discussion. Consider, first, questions of credit, responsibility, and quality. It is axiomatic that in the academic world, authority is all: we need to know who was responsible for what, and we need assurance that it is good. This serves a dual purpose: we may trust work done by a scholar we have learnt to trust; in turn, we may learn to trust a scholar from the work he or she does. From the same roundabout, scholars receive credit for the work they do, and this is a currency they can trade for promotions, conference invitations, and credibility in grant applications. But vast areas of the Internet, like the world at large, are careless about these matters which weigh so heavily in the academy. Accordingly, as scholars, we have to make a special effort both to secure proper credit for digital work and to ensure that quality-control systems remain in place. The prescriptions Bagnall sets forward to meet these needs are well described. I found this part of his paper particularly useful: we (and many others) are struggling with these issues, and he and his group are further ahead than we are with thinking these issues through and seeking to implement solutions based on the Pleiades system. I have seen prototypes of the editing system ("Son of Suda online") and it promises to do just what it says it will. That is: to provide a controlled editing environment, which does the necessary housekeeping to maintain a record of who did what and when to underpin accreditation, responsibility and quality statements. Here is our first rule, and a very uncontroversial one: *scholarly communities in the digital world require rigorous and complete declarations of credit, responsibility and quality.*

Consider, next, the matter of control. In the traditional world three things are yoked together: authorship of materials; the assurance given by the author of the quality of the materials (often supported by peer review); and the control of those materials, in the form of the right to authorize publication of the material. At the center of this nexus is the academic author. He or she creates the work, warrants its quality, and decides who may publish it, where and when.

This model, which has worked so well for the academic world for centuries, is a recipe for disaster in the

digital world. Bagnall describes very well the conflict between the “ist mein” mentality and collaboration in an open-access environment: between the imperatives of academic ownership and the open-ended partnerships characteristic of, and enabled by, the digital medium. I can speak with great feeling on this. As many of you know, for the last twenty years I have been pursuing for the manuscripts of the *Canterbury Tales* what Bagnall and his collaborators are doing for papyrology. The project has achieved much: most visibly, seven CDs published between 1996 and 2006. We have three more CD-ROMs ready to publish and we have a mass of other work contributed by project partners that we wish to continue working on. Most of all, we have complete transcripts of around 40 percent of all the manuscripts of the *Tales* ready to go online. This represents the work of some twenty or more scholars, at various levels, and a large amount of funding. But for the last five years, the project has been paralyzed. We cannot publish the CD-ROMs we have ready; we cannot publish online all the materials we have. This is because two people who worked on the transcripts of key manuscripts ten years ago have persuaded their university to withhold agreement for us to publish materials they worked on. The point at issue is not whether they or their university are right or wrong to do this: they clearly feel they have good reason for their actions, and their university supports them. The point is that the traditional model of academic ownership gives control over the work done—the right to say who may publish—to the academic who made the work. This is excellent for print publication, where the publication is the end of the work, in every sense. Once we have the book in hand, the materials used to create it (the transcripts, apparatus, piles of index cards) are of little or no interest or use to anyone. But this is most manifestly not the case for digital work. Of course, it will be useful to publish the transcripts online and on CD-ROM, just as they are, and for people to keep looking at them for years to come. But we now know that the transcripts can be far more useful than this: they can be used as the base for decades more work by other scholars, who might modify them, elaborate them, correct them, add more and more to them, republish them, as the shifting world of scholarship determines. This is the same model that prevails in the open source software world, and which has spawned the Creative Commons movement.

Accordingly, we have for the last ten years insisted that all contributors to our projects agree to the Creative Commons attribution share-alike license. The combination of this license with the established moral rights of authors affords the right mix of accreditation for the original authors with open access to all. To spell this out: there are three components of this legal framework. The first is the requirement for attribution: this mandates that every republication and reuse made of the materials must reproduce the declarations of responsibility and credit affixed to the materials. This guarantees the perpetuity of statements of credit, as required by our first rule for scholarly communities. The second component is that all publication and republication must be “share-alike”: this mandates that any republication of the materials must be on exactly the same terms as the original publication. A publisher could not, for instance, take these materials, adapt them, and then prevent anyone else republishing the adapted materials. (On the other hand, this does not forbid publishers from including the materials within a commercial site: I discuss this more below.) The third component is that of authorial moral rights. In the endless discussions among scholars about intellectual property over the last decades, moral rights has hardly figured. Yet, it is a powerful tool. In particular, moral rights permit the author to forbid inappropriate publication: we might, for instance, decide that our transcription of Codex Sinaiticus should not appear on the Wicked Pictures website.

It appears, to me, that this gives the originating scholars every right they ought to have. Indeed, there is only one right that this combination does not give the scholar who created the material. It does not give the scholar the right to say: I permit this scholar (scholar A, who is my friend) to work on and republish the materials I first created; but I do not permit this other scholar (scholar B, who is not my friend) to work on and republish the materials. This may be something we can debate: but I cannot imagine a single circumstance in which this is defensible in the academic digital world. It may be base human nature to want to use the work we have done, and think we own, to fight wars for us; as a token with which to reward our friends and punish our enemies. But the point of rules is to forbid us from doing things we might want to do, but which we all agree damage our communities.

I stress this point at length because scholars are human. And the human thing is to say: yes, all digital materials should be available free to everyone, but actually to mean: yes, everyone else’s digital material should be available free to me, but I am going to control who makes use of my material. There is no getting

around this. Open means open, and all means all. Further, as Bill Clinton so memorably failed to say: “Is” means “Is.” In our context, “Open to all” must mean: actually, really, open to all. Hence, our second rule: *digital scholarly communities must be built on open access by all, to all.*

But it is not enough to say this, as a principle we all say we subscribe to. How do we actually enable open access? Far too often, what actually happens is that scholars and project leaders think that open to all simply means that anyone can see the work on a free-to-all website somewhere. That is: you have to go through the interface provided by the scholar to get to the data. The result is that you can only see the data the way the interface permits. All too often, too, this means you cannot get to the original files themselves. Typically, the browser shows the original XML converted to HTML, with no way to access the XML. If the reader wants the XML, to work on and perhaps republish, he or she has to write to the scholar who originated the materials. The scholar may be very willing to hand these over—or might just be too busy, or might not even have access to them.

Bagnall has several pertinent arguments here. The first is the tyranny of the interface: as he points out, interfaces create project silos, with the result that the data is inaccessible to the burgeoning array of tools which other scholars might want to use on the data. It is astonishing to me that so many digital humanities sites, created often at vast expense, cannot be searched by Google. The interface locks out Google, and any other search engines, and indeed everything apart from the tools authorized by the project team. His response to this is startling: he proposes to liberate the data from the single interface, so that anyone can write a different interface: “both data and code will be fully exposed. Anyone who wants can write an independent interface.” This flies completely in the face of orthodox practice in digital projects, where the project team goes to considerable lengths to craft beautifully-fashioned interfaces—and, collaterally, decides without even thinking that there is no need to make possible any other access as the project interface does it all. Again, without even thinking it: the interface serves as a means of control. It allows scholars to pay lip service to open access (“anyone can see my site!”) while continuing the scholarly game of controlling who can do what to the materials collected in the site (“Of course, I will give permission to anyone to make use of my materials, if they ask, and...”).

This brings us to the key question of sustainability, perhaps the single most urgent issue facing us these three days of this meeting. Bagnall (in one of the few weaknesses I find in this paper) does not link the issues of control, sustainability and interfaces: but they are intimately related. It is absolutely true, as Bagnall comments, that “Control is the enemy of sustainability.” In the case of the transcripts we have made for the *Canterbury Tales*, we are acutely aware that if anyone in future has to go through the same process of negotiation that we have had to endure, for ten years now and counting, they simply will not bother. It would be quicker and more certain just to start again. If we had known, years ago, that we—we being many of us, all round the world—might not be able to publish our work on these transcripts, we would assuredly not have bothered. In turn, those original transcripts would have disappeared, quickly or slowly, as scholars turned away from them, either to work in areas unclouded by issues of control or to new materials genuinely free to all. Again, open to all means open to all.

But there is a world of difference between being really available, really accessible, really reusable, really capable of elaboration and free republication, and being so in theory only. The difference is the interface. Although Bagnall does not say this, really open data actually does depend on an interface: but an interface very different from the interfaces we have seen up to now. I agree with Bagnall that the interfaces provided by projects are the enemy: they lock away the data in silos. Worse, as the interfaces die, the data locked in them dies too. As interfaces are far the most vulnerable of any aspects of a website to decay, with bits falling off them every time a browser or operating system updates, this is a major problem. So, Bagnall is quite right to assert that we must allow anyone who wants to write an interface. But he does not spell out how this is to be done. He speaks in the sentences cited above of “data and code” being “fully exposed.” What does exposed mean? And where will the data be? And I am somewhat puzzled by the reference to “code” here (unless, of course, we are speaking of the XML encoding within or attached to the data, which makes it part of the data itself).

Here I am pleased to say: I think we are ahead of Bagnall, in developing an architecture for really open data. “Exposed” means an interface: but not an interface such as those we see everywhere. Instead, in

the architecture we are developing for the workspace for collaborative editing, the interface is metadata, so constructed as to allow intelligent navigation of the data. A full description of this lies beyond the scope of this paper: Federico Meschini and I will be presenting it as a paper at the next Digital Humanities conference in London. Briefly: Federico and I have developed an ontology of works, documents, and texts, which allows us to identify precisely, down to the level of the individual mark, exactly what texts of what parts of what works are found in just what documents, and exactly what web resources there are out there relating to those texts. Following the lead of NINES (and many others) we have implemented this ontology in OWL (Web Ontology Language) as RDF subject-predicate-object statements, as follows:

The work the Canterbury Tales contains the General Prologue, line 1

The document the Ellesmere manuscript, page 1r, contains an instance of the text of the General Prologue, line 1

The web address <http://mytranscript>¹ contains a transcript of the instance of the text of the General Prologue, line 1, as it appears on folio 1r the Ellesmere manuscript.

The web address <http://myimage>² contains an image of folio 1r of the Ellesmere manuscript

Statements such as these, retrieved (let us say) from an RDF store using SPARQL or some equivalent technology, will allow a web browser to find (for example) all pages of manuscripts containing the first line of the *Canterbury Tales*; then to find images of all these pages; and then to find transcripts of all those lines in those manuscripts, etc. I should add too that we have designed this system to be compatible with the major existing systems of cataloguing documents, works and texts, particularly the FRBR and CIDOC CRM schemes. Thus (as we have imagined it) you could find the *Canterbury Tales* and thence all the resources relating to it, down to the individual transcript of this line in this manuscript, through your online catalogue.

In this architecture, the web resources—transcripts, images, annotations—can be anywhere, and made by anyone. Now, this appears to conflict with the first rule of our community that I declared above: that we require rigorous and complete declarations of credit, responsibility and quality. Actually: it does not. The RDF system allows us to attach statements of credit, responsibility and quality to everything we make: the equivalent of the “I approve this advertisement” statement affixed to political messages. Accordingly, one could easily retrieve, and include in one’s interface, only the transcripts approved by (for example) the International Digital Project partners; or the International Greek New Testament Project, or any other body. At the same time, the system is open to materials from anyone, and one can imagine ways in which good work could be recognized and rise to the top of the sorting process, in parallel with formal academic reviewing systems. Our aim here is to unite the traditional scholarly virtues of formal structure and authority with the vigor and accessibility of the Web.

I do not assert that our scheme must be the way forward. But I do assert that some such scheme must be created if we are to have real open access, in perpetuity, to really open data. Hence my third rule: *open data on the web must be available to any form of access through intelligent metadata*. Thanks to Web 2.0 and other technologies, most of the tools and standards we need are in place. We have OWL, and multiple projects have developed experience in RDF and related technologies. I mentioned NINES; in Europe the Discovery and other projects have levered RDF into their infrastructures. We can do this.

Bagnall implies, without fully stating his reasons, that moving away from project-crafted interfaces will aid the sustainability of digital resources. I’d like to spell out, further than he does, why he is right. In our architecture, we propose that all the fundamental elements of digital data—both the data itself, and the metadata describing it—is expressed in forms readily stored within fundamental digital library systems.

¹<http://mytranscript/>

²<http://myimage/>

We are already doing this with our projects in Birmingham: we are moving the data and metadata from these into our institutional repository. Because we are able to express all our data in standard forms (as image files in TIFF and JPG; as text files in XML; as metadata in RDF; with further metadata generated automatically from the data we deposit), these are easily stored within the institutional repository. Because the institutional repository is seen as a core university service, as central to the university as email and the library catalogue, this gives the best guarantee I can imagine that our data will survive. Because, too, this is our local repository it is responsive to our particular needs, for example control over sensitive material. Bagnall points out that we must reduce costs if we are to achieve sustainability. This approach reduces the costs for fundamental storage to a level readily carried by a university. Further, linking this to the massive world of digital library software carries many benefits: as digital library software becomes ever more sophisticated, the access to and tools provided for our data in digital library stores will become ever better.

I expect by now the reader to be thinking: excellent, we can make the data sustainable. But by removing the link between data and interface we find in all existing digital projects, are we not also kicking away the ladder that allows people to get to our data? Even if the data, as we suggest, can be stored forever, cheap, it is useless if people cannot get to it. We are relying, rather heavily, on metadata to allow others to create interfaces into our data far into the future. This seems a big ask. Who will make these interfaces? Who will pay for making these interfaces?

I have two answers to these questions. The first, many of you may not like. I think there is a real role here for commercial providers. It seems to me very likely that people will pay to get good access to well-filtered data, and providers will invest in systems to give this access. Further, in many cases the providers may also hold high-value proprietary digital data that can be included in the same paid-for gateway, so enhancing the value of “free to all” data that the provider includes alongside the proprietary data. There is no conflict here with the open-to-all requirement of my second rule. The creative commons license prevents commercial providers from having exclusive access, and it prevents the provider (say) changing a few words of the original material and then claiming control of it. Indeed, I think we should welcome the prospect of the return of commercial agencies to our field. Over the last decades they have been one of the main drivers of innovation: think Chadwyck-Healey and OUP in the eighties and nineties, Brepols throughout, and Google now. They have much to contribute. Again, open to all means all. I am encouraged to see that this is already happening. A glance over the NINES and 18thConnect sites shows that many commercial publishers are already here.

The second answer you will like rather better. I think individual academics, and interested and committed individuals outside the academy, will make these interfaces, focusing on the areas of interest to them. In essence, these interfaces will be web portals, like so many already out there, only richer. The tools are already readily available (most of them) for making these portals—but we do not see them because the materials are locked in project silos. Alongside individuals, we can expect scholarly groups to make these interfaces, adding them to the websites they already maintain. Sometime soon, I’d like to see the New Chaucer Society website have a toolbar. From this you can nominate any line of any poem by Chaucer. Another click, and you can see a list of all the manuscripts that have this poem; a list of all the manuscript pages that have this line; links to all the images available on the web anywhere around the world of these pages; links to all the transcripts available on the web anywhere around the world of these lines on these pages; links to all the commentaries, glossaries, etc., available around the world for the words in these lines. Also: this should be completely dynamic; within seconds of a library in (say) Italy making available digital images of a Chaucer manuscript, the interface will discover those images through the metadata and include links to them. Furthermore, following the model of YouTube and Google and others: I should be able to drag that toolbar to my own browser, and go direct to the data anytime I like, from anywhere I want.

This is, I think, exactly what Bagnall too would like, and (if I understand it) what the IDP project is heading towards. Further, we would all like this, for every scholarly domain. And there is nothing impossible about this. So, how can we make this happen? To answer, I’d like to follow Bagnall’s lead once more. In the last pages of his paper, after outlining what he thinks needs to happen within the papyrology community, he turns to address the wider community. He asserts, rightly, that what he wants to see happen among papyrologists depends on developments outside papyrology: on developments “transcending the limited scale

of the discipline and its separateness.” He gives some precise instances of how these developments might work, in terms of shared infrastructure and cross-searching.

I think we can go further than this, on the basis of the three rules I have given above. For me, the best thing we could do in digital humanities over the next decades would be, first, to ensure that all new projects across the whole landscape conform to these rules and, second, to translate all existing projects so that they conform to these rules also. This need not cost very much money. Indeed, to the extent that it removes the need for projects to create and maintain elaborate interfaces (often one of the most expensive aspects of a project), it may lead to considerable savings. On the other hand, if we go on pouring money into creating data which then gets locked up into project interfaces, which then need yet more money to cycle through changes in computing technology, we will indeed be wasting money on a grand scale.

Who is to enforce these rules? No academic is ever going to have a problem with the first rule. We all agree that scholarly work must be rigorously credited and reviewed; we will do this without anyone requiring it of us. But this is not the case for the second and critical rule, that open to all really means open to all. Frankly, I don’t think we can trust individual academics, or even academic associations, to enforce this. The requirement for open access must be enforced by the funding agencies. In my own field, and I think Bagnall would agree in papyrology, it should be absolutely mandatory for any funded project creating digital images or text transcripts of original materials to make these available to others under the Creative Commons Attribution Share-alike license, or similar. Further, following our third rule: make available means make available the base materials and metadata to all, so that others can build their own interfaces. One could add a further rule: that this availability should be built on a credibly sustainable infrastructure, such as an institutionally maintained digital repository.

Up to now, funding agencies have been rather forgiving in their acceptance of assurances of open access and continued availability. Most of the time, they accept availability of some view of the data on some free-to-all website somewhere, together with an institutional declaration that the website will continue to be maintained, as sufficient. According to the rules I outline in this paper, this is not enough, nor even near enough. What I outline here for the humanities is already standard practice in some other disciplines. Here, for example, is the opening page of the NIH Public Access site:

U.S. Department of Health & Human Services

National Institutes of Health Public Access
The Public Access Policy ensures that the public has access to the published results of NIH funded research to help advance science and improve human health.

Home

Overview

The [NIH Public Access Policy](#) ensures that the public has access to the published results of NIH funded research. It requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funds to the digital archive [PubMed Central](#) upon acceptance for publication. To help advance science and improve human health, the Policy requires that these papers are accessible to the public on PubMed Central no later than 12 months after publication.

How to Comply

All of your papers that fall under the NIH Public Access Policy, whether in press or in print, must include [evidence of compliance](#) in all of your NIH applications and reports.

1. Determine Applicability

2. Address Copyright

3. Submit Papers

4. Include PMCID in Citations

Policy Details

Info on Journals

Training/Communications

Figure 1

I am struggling to think of a single project in the humanities that meets this standard. Even NINES and its relatives, which in many respects is very close to the model I give here, does not make available all the metadata on which it is built (or if it does, I could not find it). NINES will let you have Collex, which you can then use to make your own interfaces into the NINES metadata—but it appears the metadata itself is unavailable.

I can foresee numerous objections. Let's deal with three. First, one could assert that the PubMed Central model does not apply to us. We don't have just pdfs: we have XML files, images, and massively varied metadata. I don't think this is a valid objection: standard digital library systems can handle all this. There is work—a great deal of work—to be done on metadata standards, which is still the Wild West of our discipline. But we can do this, and enough is there already to start.

Second, one could object that funding agencies never provide, and can never provide, sufficient funding to meet the full cost both of digitization and of providing access for perpetuity. Accordingly, we need to withhold open access as I have outlined it for at least part of the digital materials we make, so that we can sustain access by charging for exclusive access to key materials. Open access and availability as I have outlined it will undermine this and destroy the business model. Large resource-holding institutions (some of them publicly funded bodies) have been eager to promote this argument, and the funding agencies have been, in my view, too ready to accept it. I propose a simple experiment. Let the funding agencies, for a period at least, mandate that funding for digital projects must follow the open-access model I here outline. There will certainly be many institutions which can work to the conditions I outline. When funding starts to flow to these institutions, other institutions will revise their viewpoint.

Third, most substantively: I am proposing, across the whole humanities, a shift to a model of access that

has (so far as I know) not been implemented fully in even one project—not even, yet, in any of my own. Further, success depends on excellent metadata. As I observe above, metadata remains the most unruly area of our work, with the freedom with which one can create, for example, RDF implementations leading to a proliferation of competing ontologies. Also, large-scale web-based systems for searching the billions of metadata records we will have are comparatively new. One might also contest my advocacy of RDF as the best way forward, and I think we should expect that we need also to handle other metadata formats, such as Topic Maps and ISOCat. But if no-one is quite doing yet what I describe, many are very close. I've mentioned the NINES and the DISCOVERY projects: these projects already have large collections of data and metadata in forms which would make possible full implementation of what I have suggested. From the other direction, we are seeing a burgeoning development of web-based tools for analysis, comparison and visualization: thus the TAPoR suite in Canada, the TextGrid tools in Germany, and many more on websites everywhere. These tools need more and better intelligent data available to them. We have the data, but it is not smart enough yet for the tools to find it, and it is too often locked away. We can bridge this gap. Finally, I do not propose the nuclear option: that funding agencies and other bodies should instantly mandate that all projects henceforward should follow these rules. Rather, let us see a few projects implement this fully. If I am right, the benefits will be so manifest that others will follow.

I see now that my response is much longer than the paper to which it is responding. Perhaps that is the best tribute I could pay to Bagnall's paper, and to the many years of remarkable work in the papyrology projects which lie behind his paper. If my paper can help us learn from the best Bagnall and the papyrologists can teach us, it will have served its purpose. If the conference participants can take something from this forward into their own work, the conference will have served its purpose.