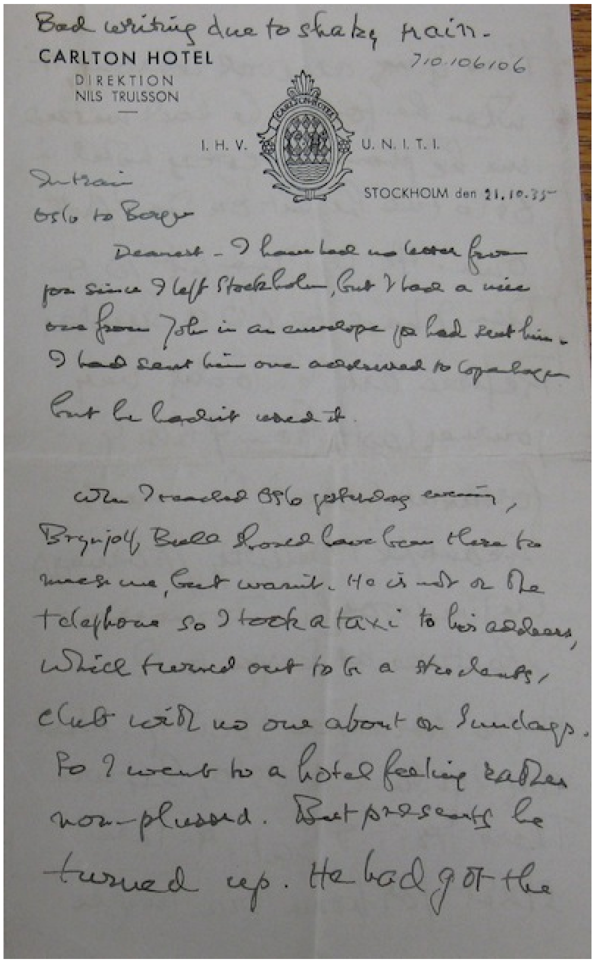
The Collected Letters of Bertrand Russell

McMaster University

Project Overview

| Overview | Details | Goals | Stages | Technical | Contact | BRRC |

Bertrand Russell to Patricia Spence - October 21, 1935



Bad writing due to shaky train.

CARLTON HOTEL
DIREKTION
NILS TRULSSON

710.106106

I. H. V. U. N. I. T. I.

In train
Oslo to Bergen

STOCKHOLM den 21.10.35

Dearest – I have had no letter from you since I left Stockholm, but I had a nice one from John in an envelope you had sent him. I had sent him one addressed to Copenhagen but he hadn't used it.

When I reached Oslo yesterday evening, Brynjolf Bull should have been there to meet me, but wasn't. He is not on the telephone so I took a taxi to his address, which turned out to be a students' club with no one about on Sundays. So I went to a hotel feeling rather non-plussed. But presently he turned up. He had got the

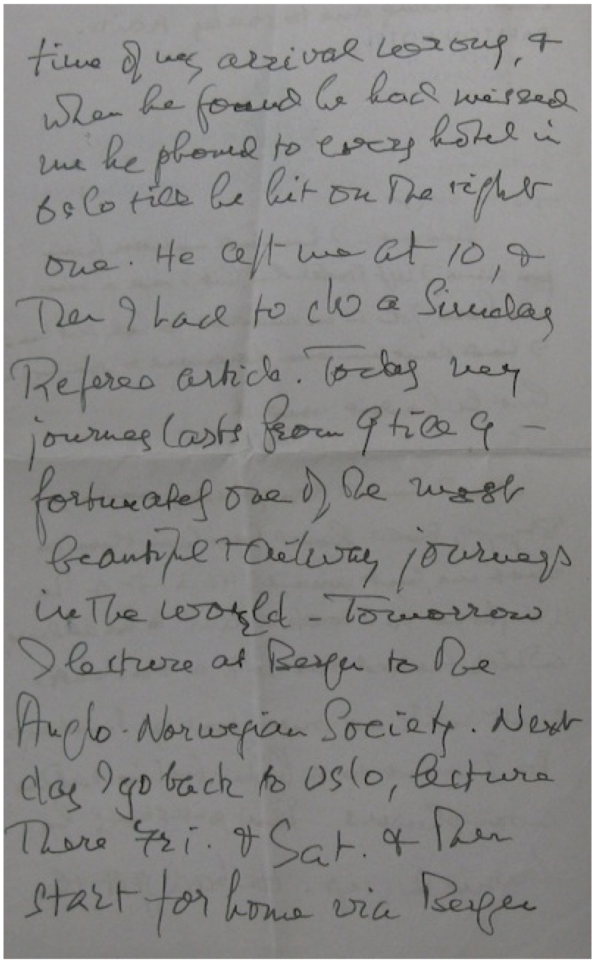
Monday, 21 October 1935
[In train Oslo to Bergen]
[Bad writing due to shaky train]

Dearest – I have had no letter from
you since I left Stockholm, but I had a nice
one from John in an envelope you had sent him.
I had sent him one addressed to Copenhagen
but he hadn't used it.

When I reached Oslo yesterday evening,
Brynjulf Bull [1] should have been there to
meet me, but wasn't. He is not on the
telephone, so I took a taxi to his address,
which turned out to be a students'
club with no one about on Sundays,
so I went to a hotel feeling rather
non-plussed. But presently he
turned up. He had got the

[1] Brynjulf Friis Bull (1906-1993), Norwegian lawyer, served as mayor of Oslo on three occasions in the 1950s and 1960s.

time of my arrival wrong, and
when he had found he had missed
me he phoned to every hotel in
Oslo till he hit on the right
one. He left me at 10, and
then I had to do a Sunday
Referee article.[2] Today my
journey lasts from 9 till 9 -
fortunately one of the most
beautiful railway journeys
in the world. Tomorrow
I lecture at Bergen to the
Anglo-Norwegian Society. Next



time of my arrival wrong, & when he found he had missed me he phoned to every hotel in Oslo till he hit on the right one. He left me at 10, & then I had to do a Sunday Referee article. Today my journey lasts from 9 till 9 – fortunately one of the most beautiful railway journeys in the world – Tomorrow I lecture at Bergen to the Anglo-Norwegian Society. Next day I go back to Oslo, lecture there Fri. & Sat. & then start for home via Bergen

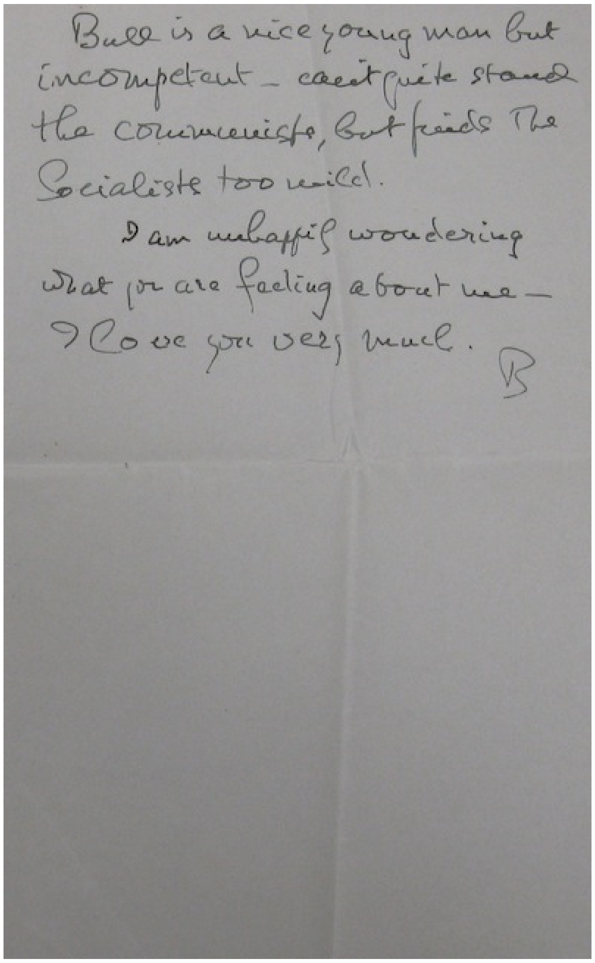
day I go back to Oslo, lecture there Fri. and Sat. and then start for home via Bergen.

Bull is a nice young man but incompetent – can't quite stand the communists, but finds the socialists too mild.

I am unhappily wondering what you are feeling about me. I love you very much.

B

[2] Probably "In Lands Where Slums and Wars Are Unknown", The Sunday Referee, 27 October 1935, p. 18; Papers 21, 60–2. This article provided an overview of Russell's impressions of Scandinavia to that point of his lecture tour.



Bull is a nice young man but incompetent – can't quite stand the communists, but finds the Socialists too mild.

I am unhappily wondering what you are feeling about me – I love you very much.

B

Overview

Russell was one of the twentieth century's great letter writers. The range of his interests, his multifarious involvement with the world's affairs, his ceaselessly perceptive intelligence, and his apparently effortless ability to translate his thoughts and feelings into words make his letters a continual source of delight and interest. He wrote to people in all walks of life and on every topic under the sun. There are many important letters about his work and many interesting political letters, including dozens to world leaders. The letters are a hugely important resource for philosophers and historians and for those interested in almost any aspect of twentieth century culture and politics.

Russell devoted some part of each day to letter writing and even in his nineties was dictating twenty or more letters at a sitting. The Russell Archives now houses between forty and fifty thousand of his letters and new ones are acquired every year. A print edition of such an immense body of work is out of the question, but current technology makes an on-line electronic edition feasible.

The edition we are planning will include a searchable digitized text of each letter. Each letter will be fully annotated to identify the recipient and the background to the letter as well as to identify Russell's allusions to persons, places, and events within it and to comment on problematic aspects of the text. Users of the edition will be able to suggest by e-mail additional information for annotations and the whole edition will not only grow continually by the addition of new letters, but will be subject to continual revision as fresh information is obtained. (The edition will include a record of such revisions so that repeat users will be able to see whether, and how, the treatment of a letter has been changed.) The edition will also include a scanned image of the original, so that alterations and doubtful readings can be seen. The letters will also be linked to enable readers to follow a particular correspondence, to read all letters over a given time period, or to search for letters on a particular topic. The edition will conform to the standards of the Text Encoding Initiative. For each letter the edition will identify the relevant documents in the Bertrand Russell Archives, explain the choice of copytext where more than one version of a letter is available, and identify the location (if known) of the original signed top-copy if it is not in the Russell Archives. In short, the edition will have all the amenities of a scholarly print edition together with some that are not possible in a print edition.

Project Details

More than forty thousand letters will be digitized, transcribed, and annotated, ultimately to be published as an electronic edition. To manage this very large project we are making heavy use of custom software to reduce the editorial complexity and control the workflow. We're using the web for most tasks, allowing contributors to work virtually anywhere in the world. And web browsers with their ever improving functionality provide a rich editing environment that we can use to hide the technical details of the data formatting.

Given the size and length of the project, a major goal is to insulate the artifacts of the project – the images, transcriptions, annotations, metadata, and other data – from technological change. The software to gather and to disseminate the data will, and should, change, but the underlying data must remain independent of any specific program.

Goals

Editorial Complexity

Fundamentally important is that nothing technical should distract the editorial work. We can't require that editors learn XML, much less the TEI tagset, or on which drive and with which filename incarnation to store a given version of a letter. The underlying storage details -- where and how the letters are stored, and how the letters and their annotations are digitally encoded -- should be hidden from the (human) editors. We've therefore developed an electronic document management and annotation system to hide the technical complexity.

Workflow

Manually managing the processing of 40 thousand letters would slow the project to a crawl, and introduce significant fallibility. We instead use software to automatically track and queue the stages through which every letter must run. Letters (images of letters) are presented over the web to annotators, transcribers, and proofreaders, in proper order. Products of these tasks: images, transcription and editorial annoations, are all made available automatically as they are finished and approved. Similarly, all new text is indexed and made available through the search interface the moment it is approved. Problematic letters are queued for review.

Public Interface

The editorial work will take decades. We nevertheless will provide, at the outset, a public web site for the letters that will display as much of the letters material as possible, as it becomes available, including images, transcription, and annotation. The site will provide well organized access to the letters, including full text search of the transcriptions and annotations of the letters, search by correspondent, by category, by date, by originating location, by form of text, and by archival class. Individual letters display images alongside their transcriptions, with annotations appearing as hyperlinked popups or hyperlinked footnotes. We have a very preliminary version of the site, but no public access until a policy for access is confirmed.

Technological Independence

We'd like, at any given moment, to remain independant of specific programs, data formats, databases, workflow systems, document management systems, and so on. Most importantly, the data should remain easily accessible and clearly organized. And so, the artifacts of any given stage of the workflow should be in a format flexible enough to:

a. allow adaptation to emerging technologies
b. ensure long term preservation

To our minds the safest option is the standard computer file system. One directory for each letter, containing: a plain text file for the transcription, the images of the letter, metadata stored in a standard format (OAI-PMH), and annotations stored as plain text files containing character offsets into the transcriptions, or if annotating images, pixel offsets into the images. The OAI-PMH files will be created from a catalogue of the letters that has been gathered over a decade.

The format for storing the annotations was a difficult decision. We initially adopted TEI embedded in the transcriptions, but in the end we've decided – in particular to allow overlapping annotations – to use character offsets to store annotations as separate files from the plain text files of the letter transcriptions. There will likely be one file for each type of annotation: references to dates, people, places, events, and finally editorial annotations.

Potentially at odds with our desire for technological independence is the desire to capitalize on new technology to quickly publish what for a year or two might be, by virtue of its novelty, a very attractive interface. In other words, to garner interest through the novelty. If, however, we are able to maintain a flexible format, we should still be able to, with judicious consideration, take advantage of new technology.

Transient Knowledge

We'd like to capture information about Russell that might otherwise go unrecorded, i.e., from the public, possibly those with direct knowledge of Russell or his letters or of the content of this letters. Someone may be able to, for example, point out that the hotel from which Russell was believed to have written a letter, was torn down before the date of the letter. The editorial annotation system described above allows editors to contribute to the project, but we'd additionally like to solicit as much information as possible from the general public. We therefore are building a system to gather comments over the web for letters.

Immediate Use

We try to make as much of the letters (images, searchable metadata and transcriptions) available to Russell researchers as quickly as possible, as soon as the information becomes available. New transcriptions should, for example be immediately indexed and made available through the search interface.

Stages

There are several relatively discrete stages. Each stage generates publishable artifacts, e.g., images, transcriptions, annotations. The stages run concurrently rather than consecutively. We won't, for example, wait for all transcriptions to be complete before beginning to annotate letters that have been transcribed. The stages:

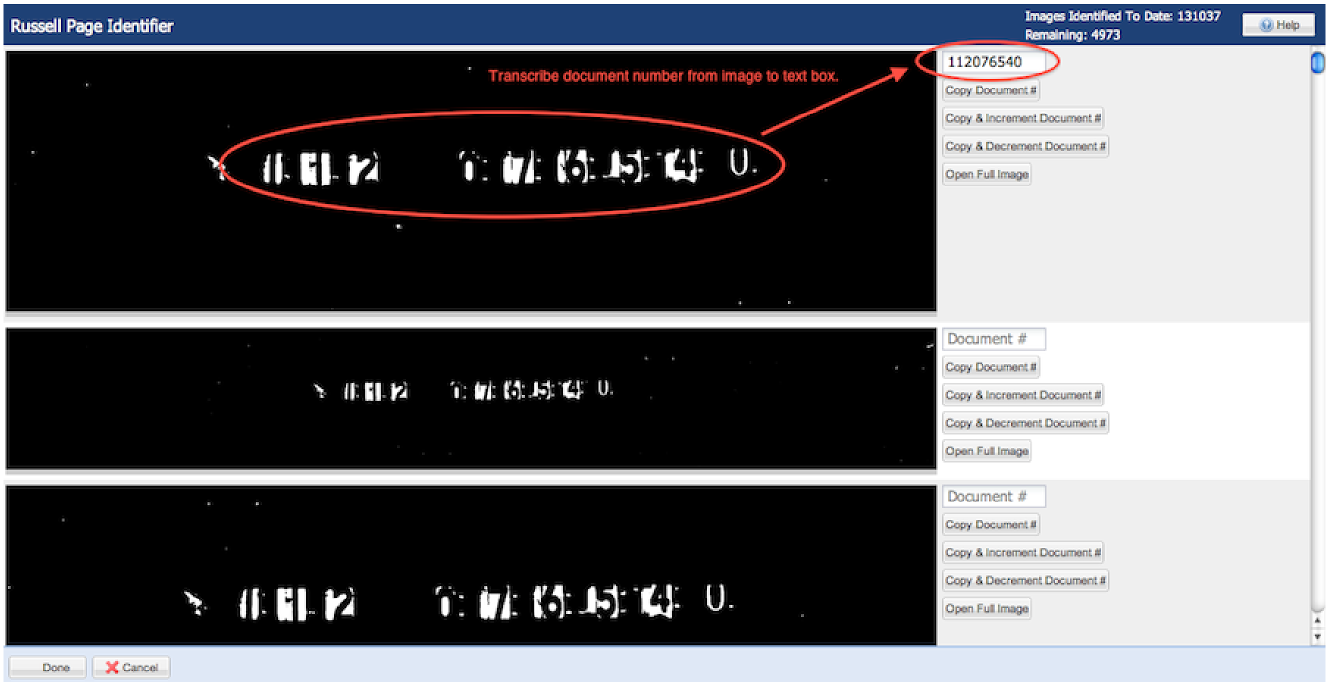
Image identification.

Over a hundred thousand digitized images made from microfilm copies of the letters had to be linked to an existing database of the letters called BRACERS. The first step in the linking was to transcribe archival numbers from the digitized microfilm images of the letters, initially painstakingly image by image, as part of a cleanup process in which the images were cropped for presentation. It quickly became clear this could take several years.

Instead, we developed a javascript web application to transcribe the archival numbers much more quickly. We take advantage of the fact that all document numbers on the images are consecutive. The numbers don't, unfortunately (and this is the problem) change every image, as there is often more than one page per letter, and all pages of a letter share the same number. Transcribers are shown 200 images at a time. They transcribe the first number, but thereafter need only push a button to increment the number when it changes.

As each number is transcribed, the database catalogue is automatically checked for a matching record. The image is automatically linked to the database record if the document number is recorded. The images for a letter will then immediately appear for a letter when the letter is shown in a search result. Many database records (over half) do not, however, have the archival document number recorded. For these letters a manual lookup process is required. This process is described next.

The following image demonstrates transcribing the first document number into the first text box by hand:



Russell Page Identifier

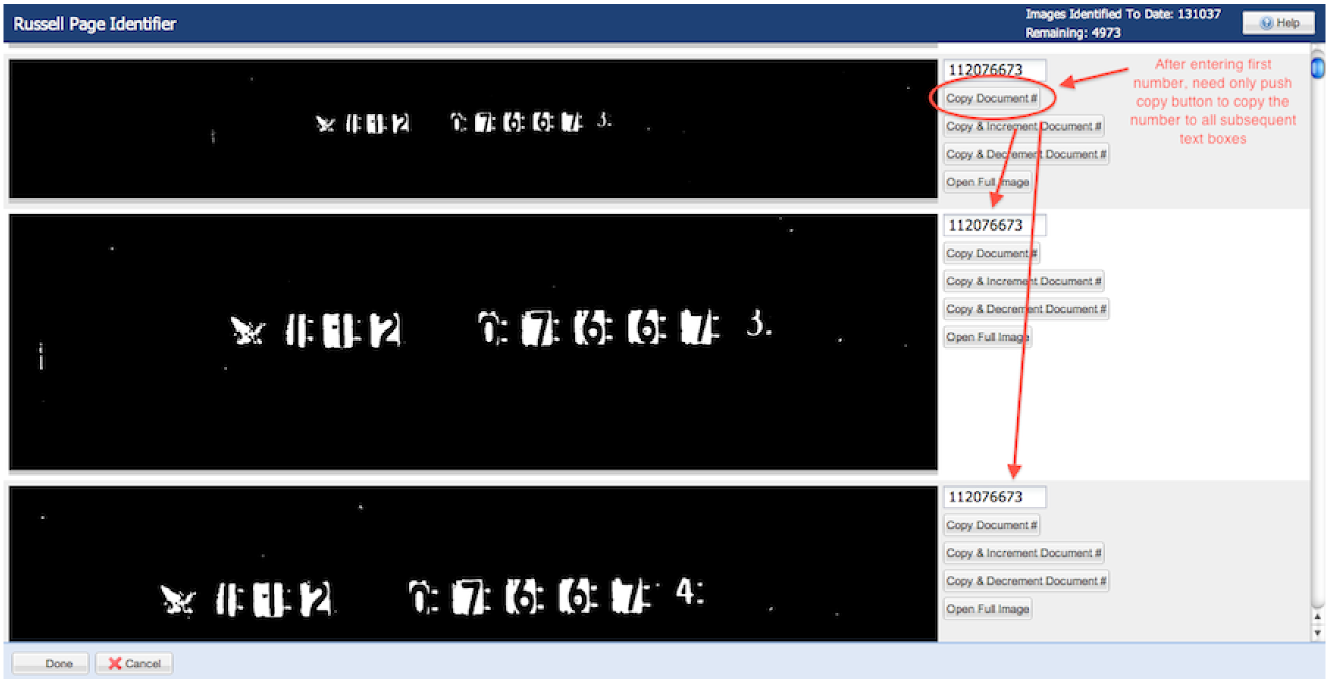
Images Identified To Date: 131037
Remaining: 4973

Transcribe document number from image to text box.

112076540

Help
Copy Document #
Copy & Increment Document #
Copy & Decrement Document #
Open Full Image
Document #
Copy Document #
Copy & Increment Document #
Copy & Decrement Document #
Open Full Image
Document #
Copy Document #
Copy & Increment Document #
Copy & Decrement Document #
Open Full Image
Done
Cancel

Thereafter, the transcriber need only push a button to copy the number in the first text box to all subsequent text boxes:



Images Identified To Date: 131037
Remaining: 4973

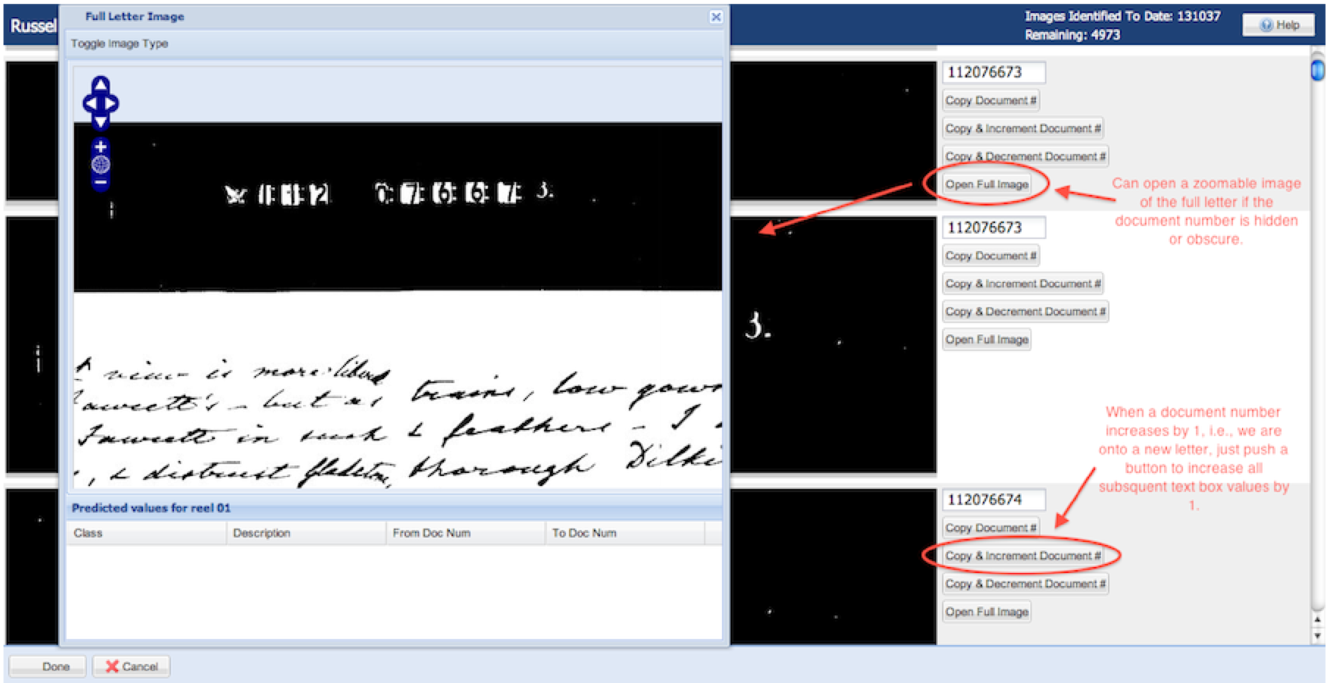
112076673

112076673

112076673

After entering first number, need only push copy button to copy the number to all subsequent text boxes

Here we can see that whenever the document number for an image increases (by 1) from the image above, the transcriber need only push a button to increment the transcribed number, and the numbers for all subsequent pages. Also, if ever a document number is hidden, e.g., we've shown the top of the page, but the document number is at the bottom, then a zoomable image of the whole letter page can opened:



Images Identified To Date: 131037
Remaining: 4973

112076673

112076673

112076674

Predicted values for reel 01

Class	Description	From Doc Num	To Doc Num

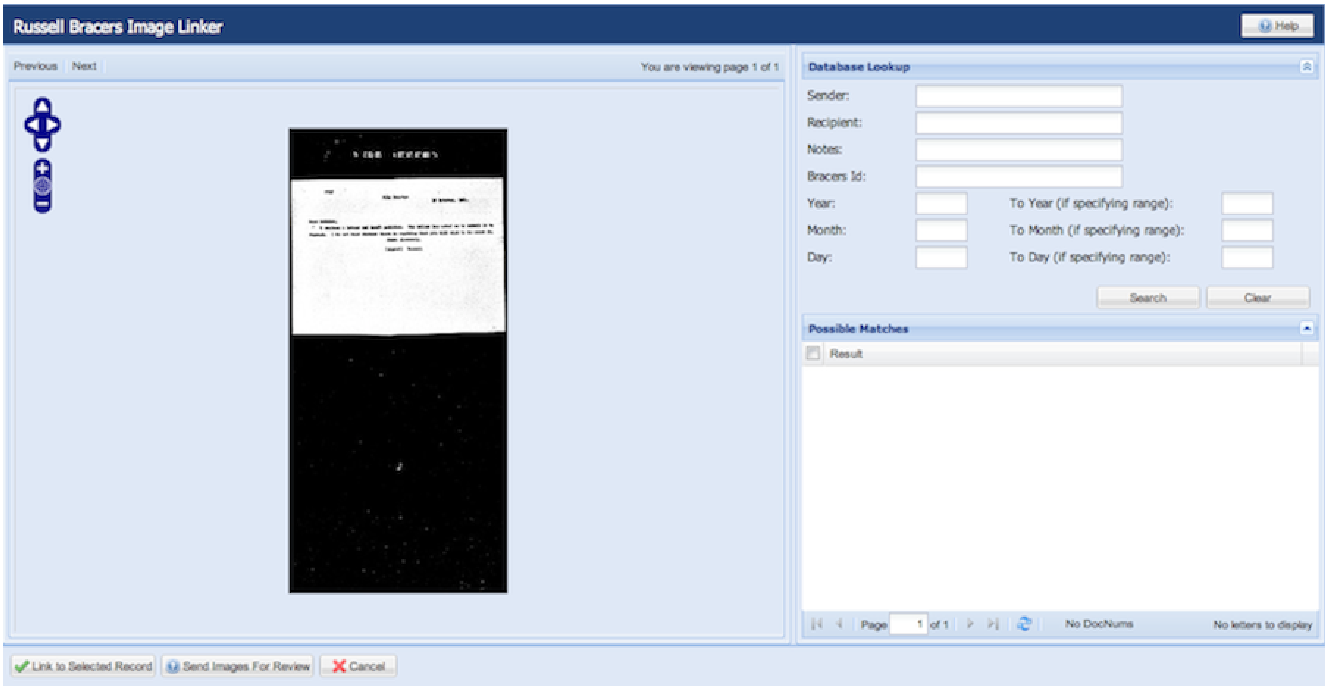
Can open a zoomable image of the full letter if the document number is hidden or obscure.

When a document number increases by 1, i.e., we are onto a new letter, just push a button to increase all subsequent text box values by 1.

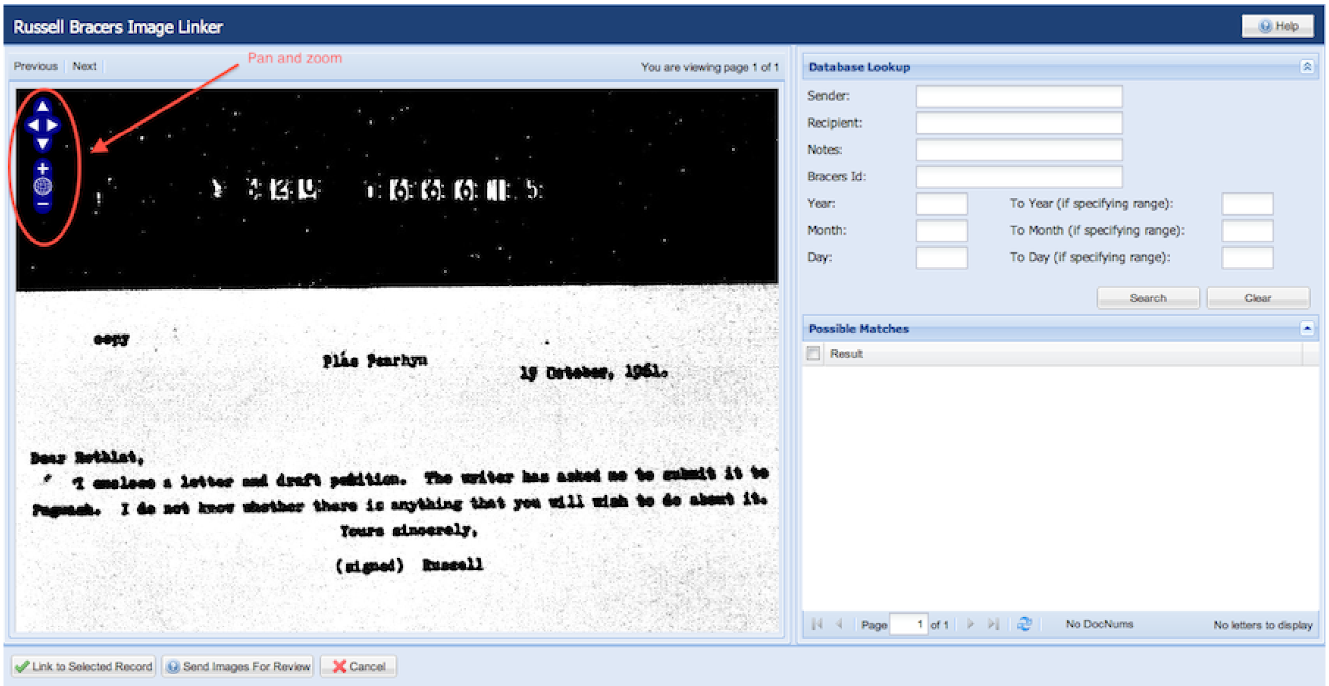
Image/catalogue association

This stage links the digitized images with their corresponding database record, for those cases where the database record didn't already contain the archival document number. For these letters, a person must manually search the database, using a web form, for the matching record in the catalogue. We've had surprisingly good success with the system, and are working very quickly through the remaining letters. As we'd hoped, the simple interface for looking up the letters, and for associating them (a single button click) has enabled us to work through over 7000 records in a matter of months.

In the following image we see the document number identifier as it appears when first opened:



And after zooming the image:



copy

Plás Penrhyn

19 October, 1961.

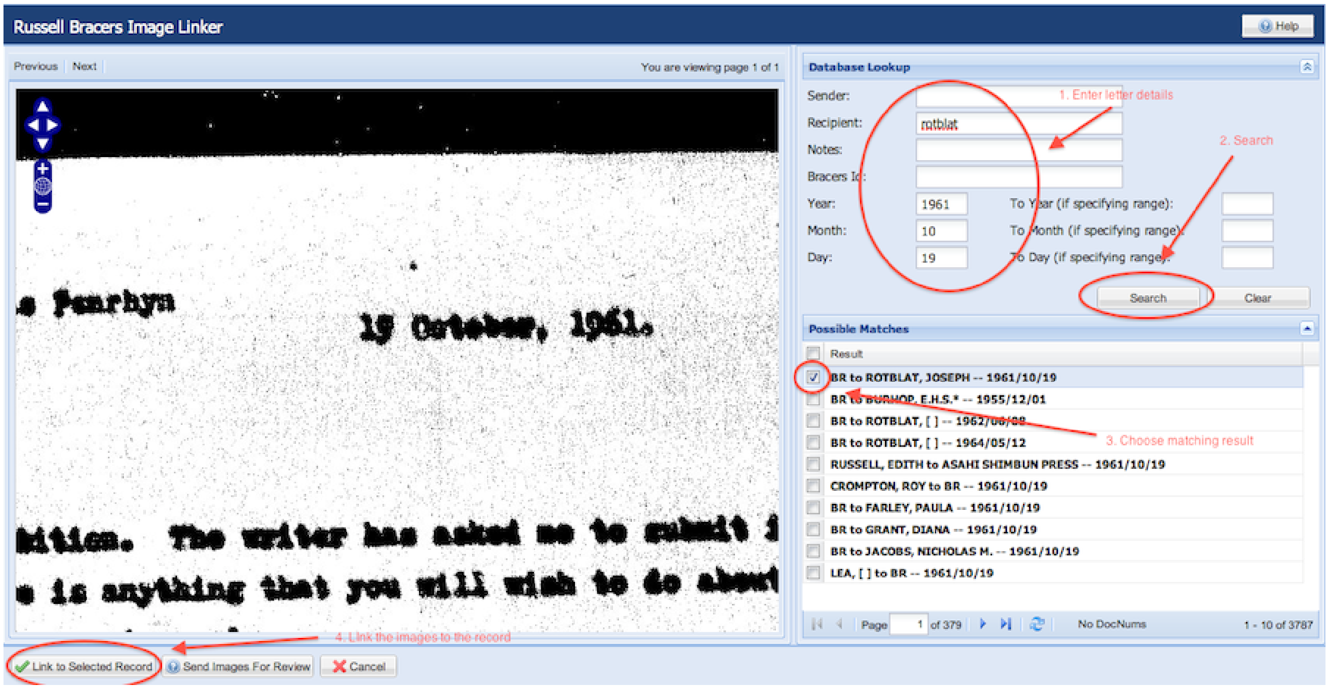
Dear Rotblat,

I enclose a letter and draft petition. The writer has asked me to submit it to Pugwash. I do not know whether there is anything that you will wish to do about it.

Yours sincerely,

(signed) Russell

And after searching for the letter and choosing the matching result:



s Penrhyn

19 October, 1961.

ition. The writer has asked me to submit i

e is anything that you will wish to do about

Sender:
Recipient: rotblat
Notes:
Bracers Id:
Year: 1961 To Year (if specifying range):
Month: 10 To Month (if specifying range):
Day: 19 To Day (if specifying range):

□ Result
☑ BR to ROTBLAT, JOSEPH -- 1961/10/19
□ BR to BURHOP, E.H.S.* -- 1955/12/01
□ BR to ROTBLAT, [ ] -- 1962/06/08
□ BR to ROTBLAT, [ ] -- 1964/05/12
□ RUSSELL, EDITH to ASAHI SHIMBUN PRESS -- 1961/10/19
□ CROMPTON, ROY to BR -- 1961/10/19
□ BR to FARLEY, PAULA -- 1961/10/19
□ BR to GRANT, DIANA -- 1961/10/19
□ BR to JACOBS, NICHOLAS M. -- 1961/10/19
□ LEA, [ ] to BR -- 1961/10/19

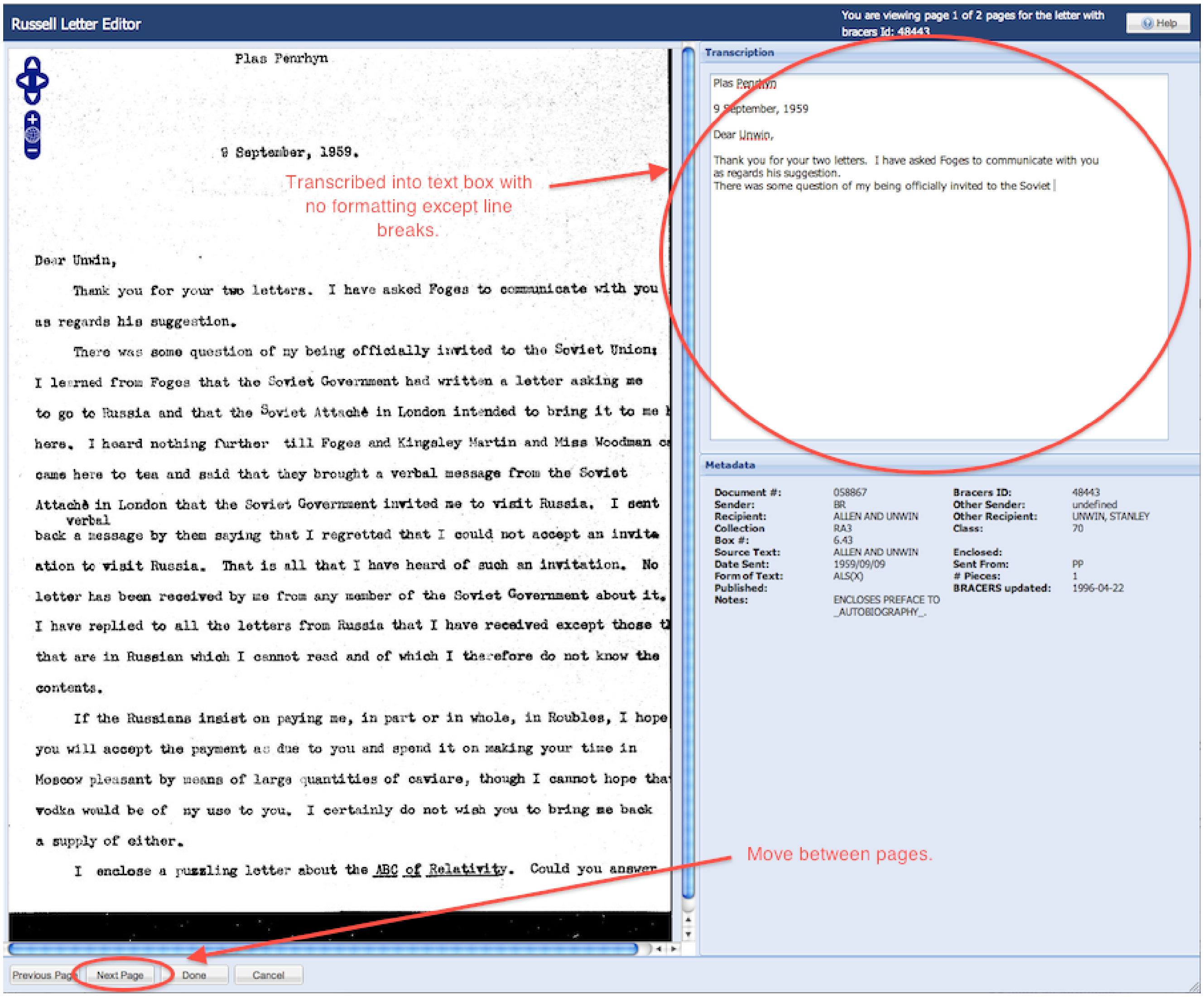
Page 1 of 379 No DocNums 1 - 10 of 3787

Transcription

The letters are transcribed manually, rather than with OCR. Many of the letters are handwritten and even those that are typed may not be amenable to OCR. Further, each letter is typically short and the form of text (handwritten, carbon, typed) varies from letter to letter. It would be difficult to automatically apply the right OCR to each letter according to the form of text. As OCR technology improves, there is always the option to incorporate an OCR engine.

The transcription stage produces plain text with no formatting. The intent is to produce searchable text as quickly as possible. Dealing with formatting issues at this point would, while producing more aesthetically pleasing text, slow down the production of searchable text. Given that images of the letters themselves will always be available for direct reading, we've chosen to defer formatting to a later stage.

This image shows the simple interface for transcribing the plain text from the letter, with no formatting other than line breaks, to the text box on the right. We see the first page of the letter.



Plas Penrhyn

9 September, 1959.

Transcribed into text box with no formatting except line breaks.

Dear Unwin,

Thank you for your two letters. I have asked Foges to communicate with you
as regards his suggestion.
There was some question of my being officially invited to the Soviet Union;
I learned from Foges that the Soviet Government had written a letter asking me
to go to Russia and that the Soviet Attaché in London intended to bring it to me b
here. I heard nothing further till Foges and Kingsley Martin and Miss Woodman c
came here to tea and said that they brought a verbal message from the Soviet
Attaché in London that the Soviet Government invited me to visit Russia. I sent
verbal
back a message by them saying that I regretted that I could not accept an invite
ation to visit Russia. That is all that I have heard of such an invitation. No
letter has been received by me from any member of the Soviet Government about it.
I have replied to all the letters from Russia that I have received except those t
that are in Russian which I cannot read and of which I therefore do not know the
contents.
If the Russians insist on paying me, in part or in whole, in Roubles, I hope
you will accept the payment as due to you and spend it on making your time in
Moscow pleasant by means of large quantities of caviare, though I cannot hope tha
vodka would be of ny use to you. I certainly do not wish you to bring me back
a supply of either.
I enclose a puzzling letter about the _ABC of Relativity_. Could you answer

Plas Penrhyn

9 September, 1959

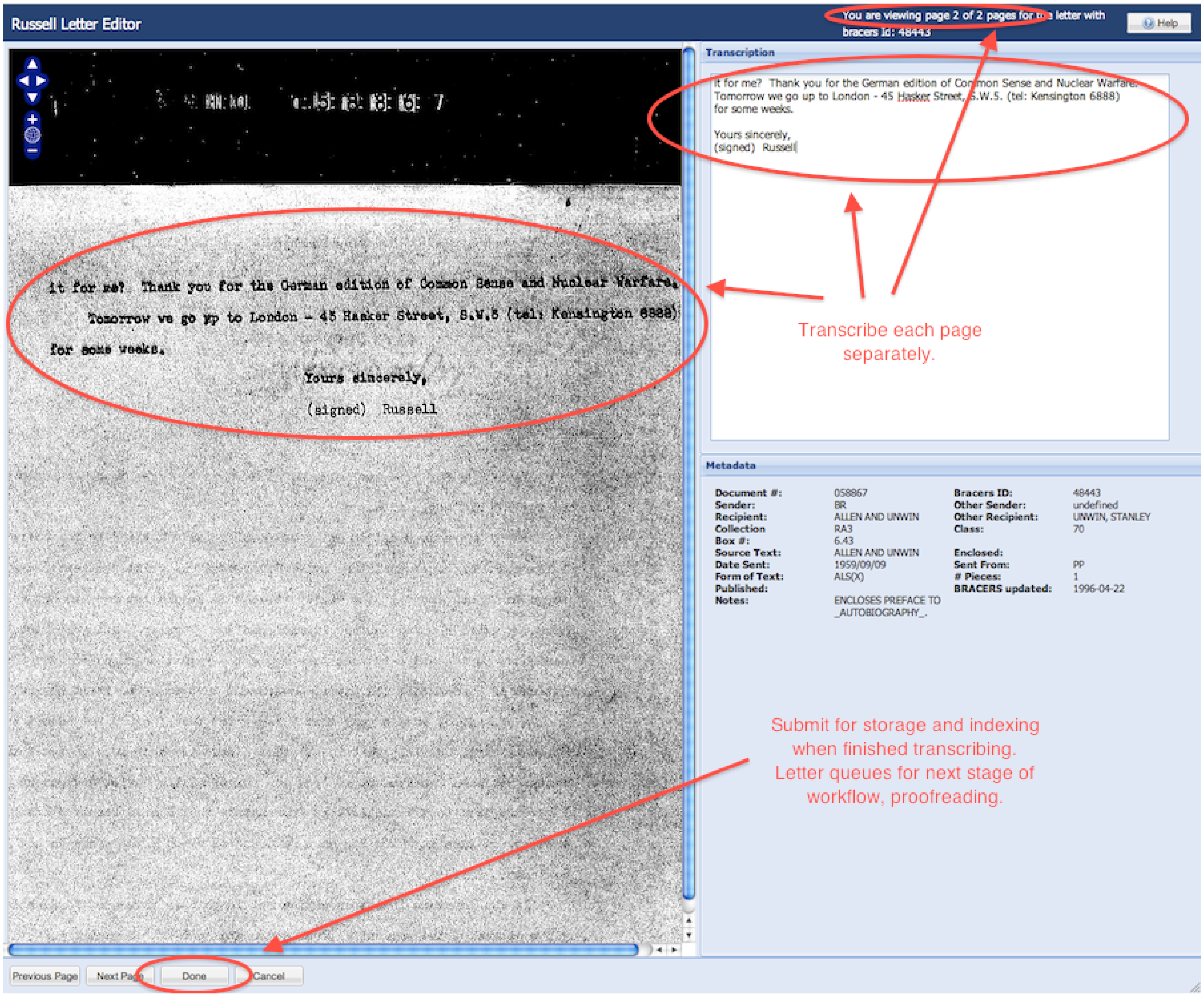
Dear Unwin,

Thank you for your two letters. I have asked Foges to communicate with you
as regards his suggestion.
There was some question of my being officially invited to the Soviet

Document #:	058867	**Bracers ID:**	48443
Sender:	BR	**Other Sender:**	undefined
Recipient:	ALLEN AND UNWIN	**Other Recipient:**	UNWIN, STANLEY
Collection	RA3	**Class:**	70
Box #:	6.43		
Source Text:	ALLEN AND UNWIN	**Enclosed:**	
Date Sent:	1959/09/09	**Sent From:**	PP
Form of Text:	ALS(X)	**# Pieces:**	1
Published:		**BRACERS updated:**	1996-04-22
Notes:	ENCLOSES PREFACE TO _AUTOBIOGRAPHY_.		

Move between pages.

This image shows the second page of the letter. Each page of the letter is entered separately. The text box on the right always shows only the text for the given page. In other words, if we move to page 2 after transcribing page 1, we'll get a blank text box in which to type page 2. If we then move back to page 1, the text box will be replaced with the text that we'd earlier entered for page 1. And if we then move back to page 2 we'll again get the text for page 2 only.



You are viewing page 2 of 2 pages for the letter with bracers Id: 48443

It for me? Thank you for the German edition of Common Sense and Nuclear Warfare.
Tomorrow we go up to London - 45 Hasker Street, S.W.5. (tel: Kensington 6888)
for some weeks.

Yours sincerely,
(signed) Russell

Document #:	058867	**Bracers ID:**	48443
Sender:	BR	**Other Sender:**	undefined
Recipient:	ALLEN AND UNWIN	**Other Recipient:**	UNWIN, STANLEY
Collection	RA3	**Class:**	70
Box #:	6.43		
Source Text:	ALLEN AND UNWIN	**Enclosed:**	
Date Sent:	1959/09/09	**Sent From:**	PP
Form of Text:	ALS(X)	**# Pieces:**	1
Published:		**BRACERS updated:**	1996-04-22
Notes:	ENCLOSES PREFACE TO _AUTOBIOGRAPHY_.		

Transcribe each page separately.

Submit for storage and indexing when finished transcribing. Letter queues for next stage of workflow, proofreading.

Entity/Date Annotation

This stage is similar to editorial annotation and in many cases will likely be part of annotation. As much as possible this stage will be carried out by students. Uncertain dates or entities will be deferred to the editorial annotation stage.

Entity and date annotation is important for at least two reasons. First, vague references to entities can be clarified. Russell might, for example refer to his brilliant but depressed student, but never by name. In confirming the reference is to Wittgenstein, the context is clarified.

Second, even if a reference is clear, say Russell referred to his brilliant but depressedn student Ludwig, it would still be better to formally identify this reference as Ludwig Wittgenstein so that a search for 'Wittgenstein' would still return letters referring to Ludwig.

Dates

Dates in the text of letters will be identified and indexed. The browser interface will allow the annotater to highlight the text of a date, and right click to bring up a form in which to record the day, month, and year. On submit, the form will create an entry in an offset annotation file, marking the start and end of the date as the number of characters from the start of the letter. The date will also be indexed for search.

Entities

References to people, places, and events will be identified in the text. As with dates, the annotater will highlight the text of the reference and right click to bring up a form in which they will enter known information about the entity, e.g., first and last name of a person. On submitting the form they'll be presented with potential matches in the entity database. They'll choose the correct match if it appears, which will create an entry in an offset annotation 'persons' file, with the start and end of the annotation together with the identifier for the entity. The entity will also be indexed for search. If no matching entity exists in the entities database, the annotator may create a new entry, or may defer identification of the entity to the editorial annotation stage.

Editorial Annotation

This is the most difficult part of the project, requiring a massive editorial effort that will take decades. Editors will complete, where possible, any unfinished or uncertain data and entity annotation, and will fully annotate the content of the letters. Editors will annotate the transcribed text of the letters, highlighting the section to annotate, and as with entity identification, will right click to open a form in which to enter the annotation. In many cases editorial annotations will complement entity annotations, for example, to explain why 'Thomas Jones" was referred to in a given letter. In these cases the editor will simply open the existing entity annotation and add the editorial matter. All annotations are listed to the side of the letter page and can be opened by double clicking.

We are additionally considering image annotation. There may be cases where something that can't be transcribed, a drawing for example, may require annotation. Or some element of the letter itself, a watermark or a stain, may require transcription.

Letter Association

Letters often answer other letters, or refer to other letters. We plan to record these relationships, most likely during editorial annotation, but possibly sooner. As with entity annotation, related letters will be recorded by looking up the related letter using a form. The matching letter will be selected from a list of returned results. The system will then record the related letter in the OAI-PMH metadata file, in addition to indexing the relation for end user search.

Proofreading

Proofreading stages will occur at several points in the project. Likely just after transcription and editorial annotation. Proofreaders will be shown the text to proofread, which they can edit accordingly. When satsifed a button is pressed, moving the letter into the next stage of the workflow. There will be multiple proofreads.

Commentary

We plan to invite public comments on the letters. We may have both a 'dicsussion' forum as well as an electronic means to comment on the particulars of a specific letter, perhaps to complement the editorial annotation. This isn't so much a stage in the processing of the letters, but will nevertheless require review, amendation, and approval.

Technical Choices

Background

The Collected Letters of Bertrand Russell project has been underway since 2002. An independent project to catalogue the letters, called BRACERS, has been underway since 1992. The BRACERS project likely has a few more years until all known letters are catalogued. Over the 8 years since the start of the project, technology has changed considerably, significantly easing the effort required to implement the interface. Despite the significant investement in early project solutions, it was faster and cheaper to abandon parts of early solutions in preference to newer solutions. In fact, at the outset, the web was the preferred interface, but browsers didn't yet provide adequate functionality. We therefore opted to build a client server app that could still be distributed and run anywhere in the world, communicating with the Russell servers. The downside, which didn't seem overly signficant on the surface, was that users had to download and install an application. We chose Java which offered compatability across major operating systems, Windows, OSX, and Linux in particular. The simple requirement that users download and install the software, however, was significant. Just enough of a disincentive to discourage users. Fortunately, web browsers have improved significantly, and specifically, javascript libraries have emerged that significanlty ease new development.

As mentioned earlier, there is a real danger of committing to a technology that becomes obsolete relatively quickly. We therefore focus on producing easily accessible and human readable data. Although we use specific technologies to produce the data, the data is always stored in as fundamentally basic a form as possible. For us this is files and directories. A relational database is used to simplify some processing tasks, but the underlying data is always stored in files.

For tracking workflow it is very tempting to adopt a workflow engine, for ease of setup and processing, but given the length of the Russell project -- decades -- it isn't hard to imagine that in short order we'd have to extricate data from the workflow engine for one reason or another. However, we also don't want to build our own workflow system, which would inevitably be an even worse system from which to extract data. We've opted to use a workflow system, but to in no way rely on the workflow system for any kind of data storage. Not just transcriptions or annotations data, but also state data, i.e., where a letter happens to be in the workflow. Progression through workflow stages is recorded implicitly in the location of files in directories.

For the annotation of the letters, we similarly wanted to adopt as basic, and therefore hopefully long-lasting, a model as possible. XML was our initial choice, and specifically the TEI (Text Encoding Initiative) Guidelines. TEI provides a well defined, well supported model for encoding scholarly material. TEI provides a ready springboard for getting an editorial project underway. We, however, ultimately wanted to move to an even more basic mechanism, for several reasons. First, TEI is XML based, which supports hierarchical data structures, e.g., a section occurs within a chapter, and a chapter occurs within a book. The immediate counter example is that a paragraph may cross a page. There are methods for dealing with these issues, milestone markers in the case of page breaks, for example, but the fixes aren't ideal. Second, we will ultimately want to 'layer' data over the letters. We'll have editorially generated content, e.g., annotations, that we'll want to display with the letters, but could have different types of annotations, or more importantly, annotations that may cross one another's boundaries. Overlapping annotations might be produced by a single editor, but we might well also want to layer annotations produced by different editors. And similarly, we may incorporate public commentary and annotation, which again we'd like to separate from other layers.

We could create and maintain several xml documents for each type of layer, but we'd then be duplicating the source text for each copy. We'd very much like to maintain a single copy of the source text. There are proposals for an 'offset' form of TEI that would allow for layering, but to further insulate our data against change, we'd opted for an even more basic format: character offset annotation, where any annotation is stored separately from the source document, and uses numerical offsets to refer to the annotated parts of the source document. So a sentence like "John took his dog to the Wellington park." could be annotated to indicate the location of Jackson Park as follows:

"25,35: Jackson park is in Peterborough Ontario."

The numbers 25 and 35 indicate the start and end of the word Wellington in the sentence when counting off characters from the beginning of the sentence. This type of annotation can be stored on single lines in a file separate from the actual text it annotates. The annotation file could contain one line for each annotation.

Storage

Our initial approach was to store everything in the file system. Files were organized in folders according to how far they'd progressed in the workflow. Transcribed images were stored in a 'transcribed' folder, along with their transcriptions. Once edited, the images would be moved to an 'annotated' folder. Once proofread to a 'proofread' folder. With the development of workflow and document management systems, the promise seemed to be systems that would take care of all the details: saving, versioning, organization, workflow. But, as has become clear, those systems will inevitably continue to evolve, requiring a 'port' to a new version or to a competing system. Especially difficult if you've committed to one that internalizes the data storage. Not just the obvious data like the texts or the annotations, but also data to do with the workflow: how many letters have been transcribed, which remain, which are problematic, etc. We've now come full circle and store everything in the file system, as we'd planned at the outset. We've additionally added versioning, but again using only a simple time stamp system on each file name, and a complementary versioning file that lists, in human readable format, who edited each file and when.

More than just the letter transcriptions and annotations are stored in the file system. We also store all of the records for the entities: people, places, and events. These records are stored separate from the letters data, one directory for each type of entity and one subdirectory for each entity. Again the entity data, like the letters' data, is indexed in SOLR for instant access.

Worth noting is that we in no way suffer performance issues when accessing the data. All of the data is indexed in SOLR, providing lightning fast access.

For some aspects of the workflow, we do still use a mysql database, but we continue to pull away from the database.

User Interface

As has become increasingly clear as technology has evolved, it is enormously difficult to predict technological change. Only lately, mobile devices like smart phones, Kindles, or the iPad promise to change how we access information. Making the Russell material as accessible as possible, ultimately will require adapting to changing technological and corresponding public viewing trends. For the moment, web based publication is still the most popular electronic medium. But even within the web, there are choices among html, javascript, flash, java applets, and so on. We've opted for the short term to adopt html and javascript as the most flexible, yet rich interface. In particular we are using the Sencha (formerly known as extjs) javascript library. Worth noting, the iPad doesn't support flash.

Authentication/Authorization

For the moment, as we continue with data gathering, we are using the simplest system possible, a bare bones authentication system. Users are simply added to a text file. As we approach publication we are moving toward the Java Authentication and Authorization Service (JAAS) system, which will allow us to eventually move to an LDAP or other similar enterprise level system.

Web Delivery

Java web delivery has run through a few frameworks over the years. At the moment we've opted for a bare bones system built only as a common web app. We are using the Restlet framework to map incoming Http calls AJAX calls from our html/javascript pages to the backend java classes that ultimately pull or save the data in the file system or in the workflow system. Beyond that we use standard web app authentication for the moment, during data gathering, but will move to JAAS in conjunction with LDAP or other, at publication time.

Workflow

The danger is in committing to a workflow system, even an open source system, that runs its course in a few years, dropping support, where our project will take decades. The competing danger is in building a workflow system ourselves, invariably introducing even more complexity and internal dependence than with an external system. Our approach is to use an open source workflow system, but to in no way store any data in the workflow system. In particular data about the state of the system: which letters have been transcribed for example, which are problematic and so on. All editorial data, as well as state data, is stored externally, in the file system so that the workflow system can be yanked at any moment.

Search

The SOLR search extends the Lucene search system. We have been using Lucene since the outset of the project, and have recently moved to SOLR with its RESTful interface, its support for faceting, and its general ease of use. Worth noting is that the search system is what provides fast access to the underlying data. End users are not searching through the file system for every search, but rather pulling from the indexed data in the search system.

Imaging

We are using the excellent aDORe djatoka jp2 imaging system along with an Open Layers viewer for djatoka created by Hugh Cayless. djatoka and the viewer provide us with zoom and pan allowing for closeup viewing of sometimes hard to read letters.

Contact Information

Project Director: Nicholas Griffin (ngriffin@mcmaster.ca)

Project Manager: James Chartrand (jc.chartrand@mcmaster.ca)