Leadership Fellows
PROPOSAL

Document Status: With Council

AHRC Reference: AH/N007654/1

Leadership Fellows (Open Call)

Organisation where the Fellowship would be held

Organisation	De Montfort University	Research Organisation Reference:	23/04/1564
Division or Department	School of Humanities		

Title of Proposed Research [up to 150 chars]

Shakespeare's Early Editions: Computational Methods for Textual Studies

Start Date and Duration

a. Proposed start date	01 October 2016	b. Duration of the grant (months)	18

Applicants

Role	Name	Organisation	Division or Department
Fellow	Professor Gabriel Egan	De Montfort University	School of Humanities

Objectives

List the main objectives of the proposed research in order of priority [up to 4000 chars]

1) To determine what the new computational techniques developed to distinguish authorship can tell us about the textual corruption and revision separating the various early editions--quartos (Q) and the 1623 Folio (F)--of the works of William Shakespeare, upon which all subsequent editions depend.

2) Where co-authorship, revision and textual corruption all contribute to Q/F differences for a single Shakespeare play, to discover how far the new methods can distinguish them.

3) To determine how best we can now explain Q/F differences for Shakespeare plays and so help today's editors to present them to modern readers, in the light of plausible theories of textual provenance, including hypotheses about how and why manuscripts of plays were recopied and how publishers received them.

Summary

Describe the proposed research in simple terms in a way that could be publicised to a general audience [up to 4000 chars].
Note that this summary may be published on the AHRC's website in the event that a grant is awarded

We know what William Shakespeare wrote only because in his lifetime, and shortly after it, his works appeared in printed form from various small London publishers. We have none of his manuscripts, so all modern editions of Shakespeare are based on these surviving printed editions. About half of his works appeared during his lifetime in cheap single-play editions known as quartos and in 1623 (seven years after Shakespeare's death) a large collected works edition of 36 of his plays, known as the First Folio, was published with assistance from his fellow actors. Where we have both quarto and Folio versions of a play, they are never identical. Hundreds or thousands of 'variants' ranging from single words to whole lines, speeches, and even scenes are present or absent in one or other edition, or are entirely reworded and/or placed in a different part of the play. Unlike the plays, Shakespeare's poems were well published and present far fewer editorial problems.

Despite centuries of study, we cannot satisfactorily explain the quarto/Folio (Q/F) variants. Some will be errors made in the printing of one or other early edition, or in the prior copying of the lost manuscripts from which those printings were made. Others will be the results of censorship that required the toning down of religious expressions used as swear-words. Others still will be the results of Shakespeare changing his mind and revising a play after first composing it, or his fellow actors changing it with or without his consent. Just which reason explains each variant is hard to say because their results can be similar. As readers and editors of Shakespeare we want to find out which reason explains each variant because we want to correct the printer's errors and censorship but not to undo second thoughts and other kinds of revision in order to show modern readers what Shakespeare actually wrote. Where he or his fellows revised a play, we want to see how it stood before and after the revision in order to understand the motivations for changing it.

The newest discoveries about Shakespeare's habits of writing concern co-authorship. Scholars used to believe that except for short periods at the start and end of his career, Shakespeare habitually wrote on his own, but we now know that as many as one-third of his works were co-written with other dramatists. This has been shown by multiple independent studies using computational stylistics, which measure features of a writer's style that are invisible to the naked eye but can be counted by machines. For the past three decades, prevailing theories of authorship have suggested that where two writers collaborate on a work they blend their styles--effectively imitating one another--so that it would be all but impossible to decide later who wrote each part of the resulting composite work. Computer-aided analysis has proved this to be untrue: personal traits of writing can be discerned even where writers attempt to efface them.

The proposed project will use the latest techniques in computational stylistics to study the problem of the Q/F variants. The techniques are particularly suited to (indeed, were first developed for) the discrimination of random corruption from systematic alteration. This discrimination goes to the heart of the Q/F variants problem: we want to know which differences result from mere errors in transmission and which are something else. Now that we have reliable tools to discriminate

authorial styles, and have a reasonable set of baseline style-profiles for most of Shakespeare's fellow dramatists, we ought to be able to see how far artistic revision by Shakespeare and/or his collaborators caused the differences between the early editions, which remain our only access to Shakespeare. The better we understand the Q/F differences, the better account we can give of what Shakespeare actually wrote.

Outputs

The main outputs of the research

Book (co-authored)
Datasets
Software
Website
Expertise gained through the application of research in a non-academic environment
If Other,

Ethical Information

Are there ethical implications arising from the proposed research ?   No

Provide details of what they are and how they would be addressed [up to 1000 characters]

Does the institution have a policy on good conduct in research?   Yes

Details of where the policy can be accessed

http://www.dmu.ac.uk/research/ethics-and-governance/research-integrity-and-ethics.aspx

Academic Beneficiaries

Describe who will benefit from the research [up to 4000 chars].

Other researchers will benefit in the following ways from the 12-month concentrated research phase of the project:

* The field of textual studies will gain a better understanding of how the early editions of Shakespeare were put together (including the provenances of the underlying manuscripts), which will enable editors to make better new editions of his works.

* Critics, editors, and literary historians working on Shakespeare will gain a more accurate knowledge of what he and his co-authors wrote because the project will improve our ability to discriminate between errors (made by scribes and printers) and the various processes of revision that play scripts went through.

* A website will give away all the project's materials--reports on experiments, cumulative discovery reports, raw datasets, programming source code, and binary executable files--so that anyone interested in how computational methods can be applied to literary and historical texts will have a ready-made set of materials from which to start their own learning and experimentation.

* A published co-authored book entitled provisionally titled "Shakespeare's Early Editions: Their Provenance and Inter-relationships" will inform the wider scholarly community of the discoveries of the project. The preferred publishers will be Cambridge University Press (who published my last book in this field, "The Struggle for Shakespeare's Text") and Oxford University Press.

* A scholarly conference will bring together experts in the field to evaluate the state-of-the-art and identify the obstacles to

further discovery and the means by which we might overcome them.

Other researchers will benefit in the following ways from the 6-month leadership phase of the project:

* A series of Travelling Roadshows will visit Strathclyde University, Liverpool John Moores University, Loughborough University, the Bodleian Library, and another university in the south-west (to be decided) to give hands-on training to academics interested in finding out about, or enhancing their skills in, computational approaches to textual studies.

* Each institution hosting a Roadshow will send a person ('the Link') to the PI's Centre for Textual Studies at De Montfort University for a week's intensive residential bespoke training in computational methods for textual analysis. Each Link will then become the champion and facilitator for these methods within their home institution.

* Included in each Roadshow--and also running as standalone events outside of them--will be a pair of interactive public theatrical performances that illustrate the fundamental inner workings of a computer and demonstrate how computers are able to store and process texts. As well as appealing to the wider public, these will be beneficial to arts and humanities academics wanting to gain a grounding in how computers work and might be put to use in their own research.

* A 48-hour Literature Hackathon at De Montfort University will enable researchers at all levels of ability to join in collaborative experimentation on live research problems, enhancing their skills and connecting themselves into networks of computational expertise they would otherwise not encounter in their research activities.

Impact Summary

Impact Summary (please refer to the help for guidance on what to consider when completing this section) [up to 4000 chars]

Who might benefit from the concentrated research phase of the project, and how?

* Individual readers and playgoers of Shakespeare (of all ages) who want to gain a better insight into what he wrote and when, including his collaborative activities, his professional career, and what is irretrievably lost to us because of errors in transmission, will be able to do so from our published results. This impact will begin 1+ years from project end and take the form of long-lasting improvement in artistic enjoyment.

* Theatre groups who want their productions to reflect the current state of knowledge about what Shakespeare wrote and how he did it will be able to do so from our published results. The success of the London replica Globe theatres and the attempts of other companies to emulate it show that paying audiences are deeply concerned with the original conditions under which Shakespeare's artistry was developed, and care about the details of his dramatic creativity. This impact will begin 2+ years from the project end and take the form of improved performances.

* Publishers of Shakespeare editions who want to give their readers--the general public as well as specialists--the latest state of knowledge about Shakespeare's processes of authorship and how they relate to theatrical practice in his time will be able to do so from our published results. This impact will begin 2+ years from project end and take the form of books that are better able to satisfy their readers' intellectual curiosity.

Who might benefit from the leadership/dissemination phase of the project, and how?

* The 'Link' from each host institution for the Travelling Roadshow will benefit from 40 hours of bespoke training in computational methods for textual analysis while in residence at the Centre for Textual Studies at De Montfort University. This impact will begin during the project and permanently enhance the Link's abilities.

* Individuals from each host institution who wish to develop their skills in computational methods for textual analysis will be able to do so by attending Travelling Roadshow as it visits each regional centre. This impact will begin during the project

and permanently enhance the attendees' abilities.

* The host institutions for the Travelling Roadshow will benefit in having their members' skills in computational approaches to textual analysis improved, not only broadening their institutional skill- and knowledge-bases in a burgeoning area of research and teaching, but also increasing their institutional capacity to undertake projects using such methodologies. This impact will begin during the project and will form permanent institutional improvement.

* The host institutions for the Travelling Roadshow will benefit from being able to offer two public performances (produced by the PI and performed by his undergraduate students) that give the general public an insight into how computers work and how they are able to store and process texts. These peformances will help host institutions fulfil their own public engagement and outreach agendas. This impact will begin during the project and will form permanent institutional improvement.

* The general public will benefit from attending the Travelling Roadshow's two public performances on how computers work and how they are able to store and process writing. These performances are interactive: audience members will be invited to join in various hands-on activities on the stage. This impact will begin during the project and comprise a societal good of improved public understanding.

* The general public, including school groups and unaffiliated interested amateurs, will benefit from being able to attend the Literary Hackathon at De Montfort University in which hands-on training in computational methods will be applied. This impact will begin during the project and comprise a societal good of improved public understanding, including among school students.

Head of Department Statement

The Head of Department or member of the Host Institution, as applicable, will complete a statement of support as a separate attachment

Name	Dr Philip Trevaskis Cox
Organisation	De Montfort University
Division or Department	School of Humanities
Post Held	Head of Department English and Creative Writing

Summary of Resources Required for Project

Financial resources

Summary fund heading	Fund heading	Full economic Cost	AHRC contribution	% AHRC contribution
Directly Incurred	Staff	144164.00	115331.20	80
	Travel & Subsistence	43607.00	34885.60	80
	Other Costs	24195.00	19356.00	80
	Sub-total	**211966.00**	**169572.80**	
Directly Allocated	Investigators	0.00	0.00	80
	Estates Costs	15795.00	12636.00	80
	Other Directly Allocated	0.00	0.00	80
	Sub-total	**15795.00**	**12636.00**	
Indirect Costs	Indirect Costs	84251.00	67400.80	80
	Total	**312012.00**	**249609.60**	

Summary of staff effort requested

	Months
Investigator	18
Researcher	12
Technician	0
Other	0
Visiting Researcher	0
Student	0
Total	30

Other Support

Details of support sought or received from any other source for this or other research in the same field.
 Other support is not relevant to this application.

Staff

Directly Incurred Posts

Role	Name /Post Identifier	Start Date	EFFORT ON PROJECT Period on Project (months)	% of Full Time	Scale	Increment Date	Basic Starting Salary	London Allowance (£)	Super-annuation and NI (£)	Total cost on grant (£)
Fellow	Professor Gabriel Egan	01/10/2016	18	100	senior	01/09/2017	81772.50	0	21907.50	103680
Researcher	Post-Doctoral Research Associate	01/10/2016	12	100	32	01/09/2017	33242	0	7242	40484
									Total	144164

Travel and Subsistence

Destination and purpose		Total £
Within UK	Travel to/from Leicester (110 GBP) and half-day's subsistence (18 GBP) for each of two external PDRA Hiring Committee members	256
Within UK	Travel (110 GBP/person) to/from Leicester and 5 days accommodation and subsistence (715 GBP/person) for each of 5 Link persons to take up their one-week residency at the Centre for Textual Studies at DMU. Total 852 GBP per Link	4125
Within UK	Six Regional Public Performances of 2 shows: Travel (225 GBP/person) and 2 days' accomm. & subsistence (240 GBP/person) for each of 10 actors plus 325 GBP travel & 3 days' accomm. & subsistence (360 GBP) for PI. Total 5335 GBP for each of 6 towns visited	32010
Within UK	Travel to/from Leicester (110 GBP/person) and one day's accommodation (120 GBP/person) and 2 days subsistence (72 GBP/person) for each of 12 technology expert speakers/demonstrators invited to the Literature Hackathon. Total 308 GBP/person for 12 persons	3696
Within UK	Travel, accommodation, and subsistence bursaries (200 GBP each) for each of 8 post-graduate students speaking at the project-ending conference on "Computational Methods for Literary-Historical Textual Studies" at DMU	1600
Outside UK	Travel, accommodation, and subsistence bursaries (320 GBP/each) for each of 6 invited international speakers at project-ending conference on "Computational Methods for Literary-Historical Textual Studies" at DMU	1920
	Total £	43607

Other Directly Incurred Costs

Description	Total £
Creation and delivery of free-standing banners and print and digital media (postcards and USB datasticks) to establish and promote the project's brand presence at Travelling Roadshows and public performances	3195
Free food and soft drinks for attendees throughout the 48-hour Literature Hackathon at 95 GBP/head for each of 120 attendees	11400
Hiring of 10 actors to rehearse and deliver the 5 pairs of public performances of the 5 Travelling Roadshows (plus an extra pair of shows in Cambridge) @ 160 GBP per actor per pair of performances, making 10 x 6 x 160 GBP	9600
Total £	24195

Estates Costs

Amount (mandatory)	£15,795

Indirect Costs

Amount (mandatory)	£84,251

Project Partners: details of partners in the project and their contributions to the research. These contributions are in addition to resources identified above.

1	Name of partner organisation		Division or Department	Name of contact	
Centre for Computing History			Communications	Ms Elaine Symonds	
Direct contribution to project			**Indirect contribution to project**		
	Description	Value £		Description	Value £
cash			use of facilities/ equipment	Use of public performance space	1152

equipment/ materials			staff time	Advance publicity & ticketing	385
secondment of staff			other	On-the-day Security, Access Control, & Front-of-House	976
other			Sub-Total		2513
Sub-Total		0		Total Contribution	2513

2	**Name of partner organisation**		**Division or Department**	**Name of contact**	
Bodleian Libraries, University of Oxford			Digital Libraries	Ms Lucie Burgess	
Direct contribution to project			**Indirect contribution to project**		
	Description	**Value £**		**Description**	**Value £**
cash			use of facilities/ equipment	Use of venue	4200
equipment/ materials	Refreshments	1200	staff time	Technical support	800
secondment of staff	Designing promotional material	860	other		
other	Digital promotion and print run	730	Sub-Total		5000
Sub-Total		2790		Total Contribution	7790

3	**Name of partner organisation**		**Division or Department**	**Name of contact**	
University of Strathclyde			Faculty of Humanities and Social Science	Professor Richard Finlay	
Direct contribution to project			**Indirect contribution to project**		
	Description	**Value £**		**Description**	**Value £**
cash			use of facilities/ equipment	Use of venue	880
equipment/ materials	Delegate packs	600	staff time	IT and security support	849
secondment of staff	The Link	2400	other		
other	Promotion and ticketing	288	Sub-Total		1729
Sub-Total		3288		Total Contribution	5017

4	**Name of partner organisation**		**Division or Department**	**Name of contact**	
Liverpool John Moores University			Sch of Humanities and Social Science	Dr Alice Cordelia Ferrebe	
Direct contribution to project			**Indirect contribution to project**		
	Description	**Value £**		**Description**	**Value £**
cash			use of facilities/ equipment	Use of venue	569
equipment/ materials			staff time		
secondment of staff	The Link	5644	other		
other	Promotion	783	Sub-Total		569

Sub-Total		6427			Total Contribution	6996

5	**Name of partner organisation**		**Division or Department**	**Name of contact**		
Loughborough University			Research Office	Professor Steven Rothberg		
Direct contribution to project			**Indirect contribution to project**			
	Description	**Value £**		**Description**		**Value £**
cash			use of facilities/ equipment	Use of venue (inc. tech+security support)		3055
equipment/ materials			staff time			
secondment of staff	The Link	6661	other			
other	Promotion	325	Sub-Total			3055
Sub-Total		6986		Total Contribution		10041

Total Contribution from all Project partners   £32357

Proposal Classifications

Research Area:

Research Areas are the subject areas in which the research proposal may fall and you should select at least one of these. Once you have selected the relevant Research Area(s), please ensure that you set one as primary.

To add or remove Research Areas use the relevant link below. To set a primary area, click in the corresponding checkbox and then the Set Primary Area button that will appear.

Subject	Topic	Keyword
Drama and theatre studies	Theatre And History	
Languages and Literature	English Language & Literature [Primary]	
Linguistics	Computational Linguistics	
Linguistics	Corpus Linguistics	
Linguistics	Textual Editing & Bibliography	

Qualifier:

Qualifiers are terms that further describe the area of your research and cover aspects such as approach, time period, and geographical focus. Please ensure you complete this section if relevant.

To add or remove Qualifiers use the links below.

Type	Name
Approach	Experimental
Approach	Knowledge exchange
Approach	Technique/Method Development
Collaboration location region	UK & Ireland
Project Engagement by Sector	Academic Users
Project Engagement by Sector	General Public
Project Engagement by Sector	Student Teachers & Undergrads
Time Period	Early Modern: c.1500 - 1800

Free-text Keywords:

Free-text keywords may be used to describe the subject area of the proposal in more detail. It is particularly important that you provide these where the Research Area(s) you have selected are only defined to two levels.

To add or remove those previously added use the links below.

Classification

Is your research multidisciplinary (i.e. involves researchers from two or more different disciplines)?

✔ Yes   No

Is your research interdisciplinary (i.e. applies methods and approaches of several disciplines)?

✔ Yes   No

Justification of Resources.

Staff Time The salaries of the Principal Investigator (PI) and the Postdoctoral Research Associate (PDRA) are the principal human-resource costs.

Directly Allocated Post: The PI
The PI must not only manage the entire project (including line-managing the PDRA) but also devise, perform, and oversee the computational experiments of the 12-month research phase of the project, teaching the PDRA how to devise and perform the experiments, and co-authoring, with the PDRA, a book about the experiments called *Shakespeare's Early Editions: Their Provenance and Inter-relationship*. The project involves creating wholly new and complex experiments in computational stylistics using electronic texts of the early editions of Shakespeare. Merely sourcing the base transcriptions is not difficult, but modifying them to ensure that they are directly comparable with one another--that their provenance does not affect out outcomes--is highly specialist and painstakingly technical work.

The experiments themselves require fresh computer programming from the PI and the PDRA, and we cannot take a 'quick-and-dirty' approach to this work. We need our source code not only to work exactly how we think it works but also to be so transparently written and so clearly documented that other investigators may not only compile the code for themselves but also rewrite it to test its operation or, indeed, adapt it to their own applications. The PI has experience in this kind of programming from his SHAXICAN project and it takes a lot of time and careful attention to do properly (see <http://gabrielegan.com/shaxican>). For all this work, 100% of the PI's time is bought out for 12 months.

In the dissemination phase of the project the PI must devise and deliver 40 hours of training for each of the five 'Link' persons from each of the five partner institutions hosting the Travelling Roadshows, as well as devise and deliver the two-to-three days of teaching of each of those five Roadshows, and organize and deliver the two-day Literature Hackathon and the International Conference. For all this work, 50% the PI's time is bought out for 6 months.

Directly Incurred Post: The PDRA (Grade 32 to meet project needs)
In the 12-month research phase of the project the PDRA will have to first learn (or relearn) about the nature and provenance of the early editions of Shakespeare in order to understand our research questions, and learn the latest techniques in computational stylistics. This will require an intensive reading programme, applicants with a demonstrated ability for such rapid learning (tested during interview) will be selected. The PDRA will have to write computer programs to execute tests devised by the PI while learning how to devise new tests of their own to pursue the research questions. By the end of the fourth month the PDRA will also have co-authored a book chapter with the PI--whilst also maintaining the extensive documentation of the project--and will maintain these parallel activities so that by the end of the 12 months the output will be an entire co-authored monograph. The PDRA will create the project website and continuously author content for it. For thisintensive period of work to be undertaken the PDRA's post will need to be 1.0 FTE for 12 months.

Directly Incurred Posts: 10 Actors for the Public Performances
In each of the five Travelling Roadshows there will be two public performances by 10 actors and the PI, plus an additional pair of performances at the Centre for Computing History in Cambridge. (The same 10 actors perform both shows: they will be selected from the PI's undergraduate students and will have rehearsed the performances on a voluntary basis as an extra-curricula part of their undergraduate degree.) The performances last 3½ hours each (including 2 hour set up and break-down) and the actors are paid at Equity Grade 1 minimum rate of £9.28 per hour plus Equity's recommended away-from-base payment of £95 per pair of performances (= £160 per actor per pair of shows) for the six pairs of performance (five Roadshows plus Cambridge show). This staff cost comes to **10 x 6 x £160 = £9,600**.

1

Directly Incurred Non-Staff Costs Aside from the recruitment costs for PDRA (advertising the post, paying expenses of candidates and external invitees to the hiring committee) there are during the during the 12-month research phase of the project no significant directly incurred costs. The hardware and software licences needed for the experiments are already available within the De Montfort University's (DMU) Centre for Textual Studies (CTS) and the datasets are being donated free-of-charge and without encumbrance by the Internet Shakespeare Editions project. During the 6-month dissemination phase, the directly incurred costs break down as follows:

Training the Link Each of the five Travelling Roadshow host institutions will send to the CTS a Link person for 40 hours (one week) of intensive training in computational stylistics. The Link's time is paid for by the host institution and the project will cover their travel to and from DMU in Leicester and their subsistence and accommodation for 5 days, projected at £825 for each of the five Links. **Total 5 x £825 = £4,125**

Delivering the Travelling Roadshows The host institutions will provide as in-kind contributions the venue, logistics--including public liability insurance and performance licence--and promotional costs for each of the five Travelling Roadshows, as detailed in their Letters of Support. The supplement the host's publicity, the Roadshow will bring its own free-standing banners and print and digital media (postcards and USB datasticks) to establish its brand presence at each event; these are costed at £3,195 for the entire run of performances. For each of the five Travelling Roadshows (plus the sixth show in Cambridge, outside of a Roadshow), 10 actors and their properties must travel from Leicester to the host institution and back, and have accommodation and subsistence for one or two days of the Travelling Roadshow, depending on how the host wishes, and is able, to schedule the public performances within the Roadshow programme. The PI will need the same travel with three days' accommodation and subsistence. The properties used in the performances, including video cameras and audio equipment, will be loaned by DMU and merely need to transported by the PI in a van costed within his travel allowance. We have costed this at £240 for two-days' subsistence and accommodation and £225 travel per actor for 10 actors plus £360 for three-days' subsistence and accommodation and £325 travel for the PI, totalling £5,335 per Travelling Roadshow (or non-Roadshow Performance in Cambridge). **Total £3,195 + (6 x £5,335) = £35,205.**

The Literature Hackathon DMU will contribute the venue (and associated logistics including publicity and small-value promotional give-aways such as tote-bags, data disks, and USB pens) and the IT resources needed for this event. Throughout the 48-hour event guest speakers from Jisc Historical Texts, ProQuest (representing Literature Online and Early English Books Online), Gale Cengage (representing Eighteenth Century Collections Online), the Text Creation Partnership, and the Oxford branch of the Text Encoding Initiative will give demonstrations of their technologies, and for each of these guest speakers the project will offer one night's accommodation, two days' subsistence, and UK-mainland travel, costed at £308 for speaker for each of 12 speakers (= £3,696). Additionally free food and soft drinks will be provided for all delegates for the duration of Hackathon at a cost of £95 per head for each of an anticipated 120 delegates (= £11,400). **Total £3,696 + £11,400 = £15,096.**

The Conference DMU will contribute the venue (and associated logistics including publicity and small-value promotional give-aways such as tote-bags, data disks, and USB pens), and the IT resources and administrative needed for this event. This three-day conference will generate income from delegates' registration fees to cover the provision of food and soft drinks for the delegates--likely to be around £50 per head, depending on catering estimate--and help to defray the travel expenses of invited speakers. To ensure the maximum impact the project also requests eight student bursaries of £200 each (= £1,600) to help post-graduates attend the conference and six overseas visitor bursaries of £320 each (= £1,920) to bring big-name overseas plenary speakers to the event. **Total £1,600 + £1,920 = £3,520**

2

Shakespeare's Early Editions: Computational Methods for Textual Studies

The enjoyment and study of Shakespeare have been revolutionized by the recent discovery that several of his plays formerly thought to be sole-authored were in fact co-written with other dramatists (reducing the size of his canon) and that plays not formerly associated with Shakespeare contain his lines, scenes, and even whole acts (enlarging his canon). These discoveries change how we think about Shakespeare's professional career, and force us to reconsider the implications of the existence, for half of his canon, of two or more early editions that differ substantially in wording and plot. This project undertakes that reconsideration using new techniques for authorship analysis.

All modern texts of Shakespeare derive from printed editions published in his lifetime and shortly thereafter, so today's readers and playgoers are heavily dependent upon the choices that editors make in transforming these to produce our modernized, critical editions. These choices are shaped by interpretations of how and why the early editions vary from one another, which remain largely subjective and aesthetic judgements. By applying the latest methods from computational stylistics, this project will generate new objective knowledge of the differences between these texts. Thus the project will put the editing and criticism of Shakespeare on a firmer footing, enabling editors to make better-informed choices about what should go in their editions. The project supports the AHRC's strategic priorities by combining traditional Shakespearian textual scholarship with cutting-edge computational methods from other fields--especially information theory and statistics--to address new questions about texts. Our answers will alter the collective scholarly view of what Shakespeare wrote, and getting this right will improve the intellectual lives of readers and playgoers nationally and across the globe.

Research Questions The cornerstone of the Shakespeare canon is the 1623 Folio containing 36 plays, including all those now thought to be solo work, together with what we now know to be collaborative plays, whose co-authors the Folio neglects to mention, and omitting several others he co-authored. For between 16 and 20 of the Folio's 36 plays we also have an earlier quarto (or in one case, octavo) edition that differs significantly from the Folio text. The uncertainty about just how many plays are affected arises because in four cases--*2 Henry VI, 3 Henry VI*, *The Taming of the Shrew*, and (most distinctly) *King John*--the quarto is so different from the Folio that we might consider it a different play. In 16 clear cases, the quarto and the Folio undoubtedly contain the same play, but with hundreds or thousands of substantive differences of phrasing and plot.

This concentrated research phase of this project is concerned with the differences between these quarto editions and the Folio itself, for which there exist no generally agreed explanations. The mid-20th-century New Bibliographers thought that the early quartos mostly reflect pre-rehearsal authorial papers and the Folio the post-rehearsal 'promptbooks' used during performance to manage the show. Critiques of the New Bibliography since the 1980s have justly undermined editorial confidence in this taxonomy, leaving today's editors of about half of Shakespeare's play with little guidance on how to choose the early edition on which to base their modernization (Egan 2010). This problem calls out for fresh empirical study. This project will provide that study by analyzing the early editions using the latest methods in computational stylistics that have enabled scholars to distinguish Shakespeare's writing from that of his collaborators. In parts the early quartos are unlike Shakespeare's writing at any stage in his career, and one explanation--popular in the 18th century and since neglected--is that these are remnants of earlier non-Shakespearian plays that he took over and revised. Complicating the picture, there are also errors of transmission separating the quarto (Q) and Folio (F) versions, as well as authorial and non-authorial revision. Thus the key questions for this research project are:

1) What can the new computational techniques developed to distinguish authorship tell us about the textual corruption and revision separating Q and F texts of Shakespeare?

2) Where co-authorship, revision and textual corruption all contribute to Q/F differences for a single Shakespeare play, how far can the new methods distinguish them?

3) In the light of plausible theories of textual provenance--including hypotheses about how and why manuscripts of plays were recopied and how publishers received them--how best can we now explain Q/F differences for Shakespeare's plays and so help editors present them to modern readers?

1

Research Context In recent years the boundaries of Shakespeare's canon have been rewritten by studies in computational stylistics. MacDonald P. Jackson (1979), David J. Lake (1979), and R. V. Holdsworth (1982) have established Middleton's substantial collaborations with Shakespeare, and Jackson invented a new attribution technique based on trigrams-in-common (2003; 2014). A rival trigrams-in-common method was developed by Brian Vickers (2009; 2010; 2012). John Burrows contributed a new way of processing the rates of frequently occurring features such as function words, called Delta (2002; 2003) and developed a new way of determining just which words are most useful to count, called Zeta (2007). Burrows's methods are embodied in the Intelligent Archive software developed by Hugh Craig and others and used to revaluate the currently contentious claims in Shakespearian authorship attribution (Craig & Kinney 2009).

Studying the frequencies of occurrences of particular words and their proximities to one another we necessarily apply Claude Shannon's principles of Information Theory (Shannon & Weaver 1949; Shannon 1951). Many recent successes in computational analysis indirectly measure the Shannon entropy of texts, avoiding direct measurement because its mathematical complexities are little understood in the humanities. The PI has successfully used direct measurement of relative Shannon entropy to compare Markov chains containing Word Adjacency Networks (WANs) that capture the word-proximity data for upwards of 100 function words across whole dramatic canons (Egan et al. 2016). This project will build on measurements of linguistic entropy for authorship attribution because it also allows us to specify the informational content of writing and to quantify its corruption in transmission (Taylor & Egan 2016).

Research Methods The WAN method (Egan et al. 2016) combines the approach of Jackson and Vickers for finding and measuring low-frequency trigrams and collocations with the approach of Burrows and Craig for weighing the rates of high-frequency function words. The method can also reproduce precisely the tests of these investigators, which between them represent the state-of-the-art in Shakespearian authorship attribution. Applying these methods to the Shakespearian Q/F differences, the project will use the newly established boundaries of his canon to differentiate co-authorship from textual corruption during transmission on one hand and from authorial revision on the other. Non-authorial revision such as Middleton's alleged adaptation of *Macbeth* and *Measure for Measure* may be treated as a form of co-authorship. Authorial revision has recently overtaken corruption in transmission as the dominant explanation for substantial Q/F differences in the plays *King Lear*, *Hamlet*, *Richard III*, and *Othello*, and is widely suspected for many more.

The measurement employed by this project, Shannon entropy, was developed to quantify textual corruption in electric telegraphs and has recently been applied to the determination of literary authorship (Shannon & Weaver 1949; Rosso, Craig & Moscato 2009). For Shakespeare we lack uncorrupted originals (authorial manuscripts) from which to measure the relative corruption of any particular printed editions. (The one minor exception, Shakespeare's contribution to *Sir Thomas More*, is too short for this method.) We can, however, measure the relative entropy between texts, quantifying how far a quarto text differs from its corresponding Folio counterpart. Kathleen O. Irace (1994) attempted to quantify these differences for the so-called 'bad' quartos, but her method was largely subjective even though she processed her results computationally. Moreover, we can rank the quartos by their fidelity to a collective Shakespearian style quantified from a subset of his canon widely considered by scholars to be reasonably uncorrupt.

Writers' styles change in measureable ways across their careers (Hope 1994; Craig 2013). The use of the *thou* for modern *you* generally declined over Shakespeare's lifetime, and although (being from rural Warwickshire) he used it more often than his London-born contemporaries, his usage declines distinctly after 1600. In his increasing use of *does* instead of *doth* (and similar constructions) Shakespeare again followed the general trend, but his rising use of the modal auxiliaries *would* and *could* and the intensifier *very* are rather more idiosyncratic traits. Because of these changing habits, when Shakespeare revised his plays long after their first composition, as many people think he did with *King Lear*, the effect is a collaboration of the older and younger selves; we will explore whether quantitative analysis can detect this.

The project, hosted at De Montfort University's (DMU) Centre for Textual Studies (CTS, of which the PI is Director), will spend one year running experiments applying the most recent methods in computational stylistics to the differences between the early quartos and Folio texts of Shakespeare's plays. The algorithms will be devised by the PI, assisted by a Post-Doctoral Research Associate (PDRA), and programmed in languages such as Python (for its Natural Language Toolkit), Perl, Prolog, C, and Lisp, with R and MATLAB for numerical processing. The final decision on languages will be made in the light of the

overlap between the appointed PDRA's knowledge and expertise and that of the PI; just as with early modern English*, tastes in programming languages strongly reflect their holders' ages and origins.

Digital texts of early editions of early modern plays will be sourced from the ProQuest database *Literature Online*, the newly released public-domain tranche of the Text Creation Partnership, and the hand-crafted scholarly transcriptions of the Internet Shakespeare Editions project (which has agreed to provide them to this project). These will be pre-processed by a variety of shell scripts created by the project to normalize them to the standards required for the various experiments. Typically, speech prefixes, stage directions, and all proper nouns are removed, although some experiments also require modernization of early modern spelling, and for those the VARD 2 package from University of Lancaster will be trialled and compared with the modernization routines embodied in the Intelligent Archive software.

The experiments will assess the current methods' ability to distinguish co-authorship from other causes of Shakespearian Q/F differences (authorial revision and corruption in transmission), using as a control the widely agreed-upon divisions of the co-authored plays *Henry VIII*, *The Two Noble Kinsmen* and *Timon of Athens*. Six of what used to be called the 'bad' quartos of Shakespeare--*The Contention of York and Lancaster / 2 Henry VI* (1594), *Richard Duke of York / 3 Henry VI* (1595), *Romeo and Juliet* (1597) , *Henry V* (1600)*, The Merry Wives of Windsor* (1602), and *Hamlet* (1603)--will be assessed for their Shannon-Jensen distance from their Folio counterparts. (A seventh, *Pericles* (1609), has no Folio counterpart.) We want know whether these quartos form a group whose members are quantifiably distant from their Folio cousins. For two of these six plays a further comparison may be made with their 'good' quarto counterparts: *Romeo and Juliet* (1599) and *Hamlet* (1604-5). The 'good' quartos will be tested to see if they too form a quantifiably alike group, as the 20th-century New Bibliographers thought. Where undoubted co-authorship co-exists with strongly suspected textual corruption--in *Pericles* (1609) and *Timon of Athens* (1623)--the aim is to distinguish these phenomena within a single play. Verification of the long-standing and currently unfashionable hypothesis that authorial self-revision underlies a number of reliable Folio-only plays, including *The Two Gentlemen of Verona*, *The Winter's Tale* and *The Tempest* will be attempted, as will quantification of the known interference of the scribe Ralph Crane who supplied the printer's copy for some Folio plays.

The research project will produce a book co-authored by the PI and PDRA on the Q/F differences in the Shakespeare canon. The book will put the project's results in the context of three centuries of research on the problem and its relationship to the textual economies of early theatre. Recent Shakespearian co-authorship studies have confirmed some of the subjective intuitions of pre-20th-century critics and long-discarded 18th-century ideas about Shakespeare taking over and revising existing plays seem newly plausible in the light of computational analysis. A website describing the project will also be created. All project materials--visualizations, algorithms, source code, compiled binaries, and project reports--will be made available under a Creative Commons BY licence to encourage the widest possible reuse.

Leadership Development Plans The study of Shakespeare's texts is archetypal to the study of literature in general, and showing what can now be done with his works provides models for what can be done with other writers' canons. This archetypal role arises mainly because Shakespeare's printed editions and professional career have acquired the largest volume of prior scholarship upon which to build new techniques. The project's leadership activities proceed from the PI's conviction that the entire field of English and Historical Studies--taken together as Textual Studies--can be transformed by new algorithmic and computational methods. C. P. Snow's diagnosis, more than half a century ago, of a divide between the cultures of science/technology and arts/humanities remains accurate and the consequences for the arts and humanities are severe. Splits are developing not only between those who do and do not understand the computational methods, but also between those who accept, or not, the methods' very validity and findings. Despite decades of scholarship in what we now call Digital Humanities, many textual scholars (literary and historical) remain to be convinced of the utility of computational methods, or else accept their utility but do not know how to start using them to enhance their own analyses.

The PI will provide the leadership to help overcome these discipline-wide obstacles. He published the first application of large-textual corpora searching to a theatre-historical problem (Egan 2001) and contributed to the discrediting of Donald W. Foster's SHAXICON methodology by collaborative open-source testing (Egan, Roth & Steggle 2002). The online version of the PI's conference paper on the use of 'regular expressions' to achieve determinacy in computational stylistics is widely cited in scholarship applying these

3

methods to Shakespeare (Egan 2005). He serves as a Technical Evaluator for the AHRC and chairs the advisory board of Jisc's Historical Texts online database, a publicly funded rival to the commercial research databases EEBO and ECCO, which (after a shaky start) it now surpasses in sophistication. He is co-editing, for publication in 2016, a volume called *Shakespearian Authorship: A Companion to the New Oxford Shakespeare* that will present the ground-breaking scholarship in computational stylistics (including his own) on which this new edition's radical reshaping of the Shakespeare canon is based. The PI's expertise in Shakespeare, textual studies, and theatre history is combined with expertise in teaching (he is a National Teaching Fellow for 2014-17) and extensive experience of service across English Studies at a national level, all of which offer the potential for discipline-changing leadership activity. To help close the gulf between computational and non-computational scholarship in English and Historical Studies, the following leadership development work will be undertaken, together with a series of public engagement activities described in the dissemination section. These activities will include public performances created by the PI as part of his undergraduate teaching, building on his four years' experience as the first Globe Education Lecturer at the replica Shakespeare's Globe theatre in London, for whom he devised and taught the first Master's degree (2000-2004) as a collaboration with King's College London.

(1) Reshaping the Research Agenda By applying the new quantitative methods to the collateral early editions of Shakespeare, the PI will put the Q/F relationships on a firmer footing than they have enjoyed since the mid-20th-century heyday of New Bibliography. In place of New Bibliography's largely intuited taxonomies of textual provenance, which have rightly been criticized for over-generalization by New Textualists since the 1980s (Egan 2010), the PI will provide new, reliable, and empirically grounded accounts of the Q/F differences and contextualize them within the current state of scholarship about the wider textual economy of early modern theatre. The result will be a new account of the likely provenance of each early edition, enabling modern editors of Shakespeare to make more fully informed decisions about which early editions to use as the bases for their own.

(2) Raising Awareness and Understanding of Computational Methods across English and Historical Studies Using the archetypal case of Shakespeare, the PI will raise the general awareness of how computational methods for textual studies work, what they have discovered, and how to individuals may employ for themselves. He has a 15-year track record of explaining to a general Shakespearian readership the developments in computational approaches via his annual reports on "Shakespeare: Editions and Textual Studies", typically between 20,000 and 35,000 words each, in *The Year's Work in English Studies* (Egan 2001-). To engage scholars it is not enough merely to explain the procedures: one must demonstrate by the creation of new knowledge that it is worth making the effort to learn about and use them. Hence the PI's concern with throwing new light on the longstanding Q/F problem in Shakespeare.

(3) Fostering Impact Most non-specialist readers and playgoers of Shakespeare assume that the texts of his works are well-established and universally agreed upon, which is far from the truth. The wording of editions of Shakespeare read for pleasure, and performed for public entertainment, are highly dependent upon their editors' assessments of the relative merits of the collateral early editions, and since the collapse of faith in the principles of New Bibliography editors have lacked guidance in this area. One recent response to this lack of guidance has been to offer both Q-based and F-based texts within the same modernized edition, which simply passes the problem on to the ill-prepared lay reader. The PI will offer a way out of the current impasse in the editing of Shakespeare leading to an improved intellectual life for his many readers, students, teachers and performers across the world by giving them better editions of his works. Additionally, a series of public engagement events (described in the section Dissemination), including a theatrical performances and an interactive demonstrations, will transfer practical knowledge about, and practical skills in, the use of computational methods to transform textual analysis.

Technical Summary The core research of this project is experiments in software using large textual corpora. The attached Technical Plan indicates how this will be achieved, including the provenance of the primary electronic texts. All algorithms, source code, and results will be published in Open Access and Open Data form on the project website and lodged in appropriate repositories. The source texts for project-- the Q and F transcriptions--will be high-quality transcriptions made by the Internet Shakespeare Editions (ISE), which has agreed to donate them in their native SGML format for use in this project. Copyright in the scholarly work in these transcriptions remains with ISE and will be acknowledged in all project outputs.

Project Management The PDRA will bring to the project specialist technical skills in Shakespeare, textual theory, and computational methods and will have career-development needs that must be explicitly addressed by the PI as line-manager. After training in the technologies of the project--including an intensive research reading programme--the PDRA will first perform experiments devised by the PI. The PDRA's ability to devise new experiments will be developed during the project, using the PI's experiments as models. The PDRA's ability to write-up research experiments and discoveries will be developed over the course of the project by co-authorship with the PI of regular project reports and the chapters of a major scholarly book. The project's Advisory Board (AB) will be six world-leading exponents of the application of computational methods to Shakespeare studies, comprising MacDonald P. Jackson, John Burrows, Hugh Craig, Doug Duhaime, and Brett D. Hirsch (all of whom have already agreed to serve) plus one more to be elected by these five. The AB's role covers the concentrated research and leadership phases of the project, specifically to 1) advise on the proposed order of business by approving/revising the experiments, 2) read and critique the regular reports on the project's experiments, looking for unconscious and methodological biases and correcting any technical omissions or mistakes, and 3) review project progress and advise on further work. The AB is qualified to assess project risks, including the likelihood of experiments failing to produce decisive results.

The key figures whose non-availability would imperil the project are the PI, the PDRA, and members of the AB. Without the PI the project is entirely infeasible and his non-availability by death at any point except near the end is an unavoidable risk that cannot be mitigated. The PDRA will be found by advertisement and candidate selection (we have no-one in mind) and so is in principle replaceable if the person becomes unavailable during the project. Replacing the PDRA would set the project back substantially and this risk will be mitigated by the scrupulous online documentation of all the project activities so that a replacement to could pick up the threads with minimal (although still significant) disruption. The expertise of the AB members is complementary but also overlapping, so loss of one or even two would not be disastrous, although it would be a set-back to the project's steering ability. This risk is mitigated by the PI having a 'reserve' list of AB members--one per specialist area of expertise.

Timetable See Workplan for detailed timings. The general outline of the experimental research phase that recurs throughout the project is that we will i) collect and prepare digital texts (the Q/F files) to work on, normalizing/regularization the variations arising from their differing provenances, ii) devise, in consultation with the AB, a series of tests to be performed on the digital texts and embody these as algorithms, iii) code the tests in various programming languages, iv) run our code on test datasets to ensure that it accurately reflects our algorithms, v) run our code on live datasets (Q and F) and collect our results, v) review our results, summarize them, and share them with the AB, vi) publish our results on the project website and decide how they should shape the next round of testing. Steps (i)-(vi) will be repeated on a monthy basis across the 12 months of the concentrated research phase. The Workplan details this rolling programme, indicating project outputs, and dissemination plans.

Dissemination The PI and PDRA will co-author a substantial research monograph of 90,000-110,000 words on the Q/F differences in Shakespeare and create a project website making available all materials, including devised software, under a CC-BY licence. Like the PI's history of editing theory and practice (Egan 2010), the book will be written in a style that makes its discoveries accessible to a general readership as well as specialists. Then follow three distinct dissemination activities to serve the leadership and impact goals:

1. Travelling Roadshow on Computational Methods in Textual Studies In a series of travelling roadshows (each lasting 2 or 3 days) the PI will lead demonstrations and hands-on training sessions in computational methods at the research centres of the Bodleian Library, the University of Strathclyde, Liverpool John Moores University, and Loughborough University (see Letters of Support), and one more to be chosen by invitation to potential partners located in the south-west of the UK which is not yet covered. Assisting the PI in each Roadshow will be a co-opted member of the host research centre (hereafter, the Link), who will liaise between the project and the host centre. Prior to the roadshow visiting each research centre, the Link for that centre--who may be a research student, a teaching tutor, a researcher, a technical project developer, a subject librarian, or suitable equivalent--will be a resident fellow at the CTS for five days during which the Link will be trained in computational methods and supported to develop their own project(s) at whatever level of attainment they possess. Where a host sends a research student, this

5

secondment will serve as a collaboration between the host's Doctoral Training Programme (DTP) and the combined DTP of the Midlands Three Cities Consortium of which DMU is a member.

The five Links will be channels for Knowledge Exchange between the project and the five host research centres and will develop those centres' capacity in computational research in textual studies after the project ends. (The PI's track-record in this area includes his leadership of a £416,000 AHRC Knowledge Transfer Fellowship in 2011-13 on the subject of "Shakespearean London Theatres", a partnership with the Victoria & Albert Museum.) The project will pay the travel, accommodation and subsistence costs of the Link while resident at the CTS and their home institution will continue to pay their salary if an employee or their studentship bursary if they receive one. The home research institution can thus treat the opportunity to be a Link on this project as a cost charged against the Continuing Professional Development or Doctoral Training Programme budget for the individual chosen. The CTS itself will not charge for the training given, although the PI's time will be charged against the grant presently being applied for.

Each Roadshow will also offer two 90-minute public performances created by the PI in which his undergraduate students--for whom these performances are part of their module on "Textual Studies Using Computers"--act as various pieces of computer hardware in a series of interactive illustrations of the principles by which machines process texts and execute algorithms for computational analysis. Using the expertise he gained at the Shakespeare's Globe theatre, London, the PI is creating these performances now for a premiere at the Electronic Textual Cultures Lab at the University of Victoria, Canada, in January 2016. A further delivery of the two public performances will take place outside of the Travelling Roadshow at the Centre for the History of Computing, a charity-run museum in Cambridge (their Letter of Support is included in this application). At least one further delivery of the performances will take place at a school in Leicester under the auspices of DMU's Square Mile project for local engagement, which will cover its (modest) costs.

1. Literature Hackathon DMU's Centre for Textual Studies will host a 48-hour Literature Hackathon on the subject of authorship attribution, dating and literary criticism by computational means aimed at all constituencies (public and professional) interested in authorship. Promotion for the event will include listings on Hackathon-Watch, Eventbrite and the usual social media, together with, for local coverage, integration within DMU's Square Mile project that connects the University with schools and social organizations across Leicester and the East Midlands. (The PI already runs a Computing Club to give local schools hands-on training within the CTS lab.) As with all Hackathons, the target audience is defined by personal interest rather than demographics, so the promotional activities (managed by a dedicated University public engagement unit) will concentrate on gaps in public awareness wherever they are found. The Hackathon itself will include a 'bake-off' competition between competing authorship-attribution methodologies and will showcase recent advances in computational literary knowledge creation. All comers will be able to try out the new methods on their own authorship questions and share their interests, practices, and discoveries. In addition to the Collaborate-to-Compete strand, the Hackathon will have first-steps sessions for those entirely new to computational methods for literary studies and intermediate sessions for those wishing to make the transition from simple explorations to serious research applications.

Attendees will get plentiful technical support--including multiple high-speed Internet connections and power sockets per person--and offline data (on USB stick and Blu-ray disks) comprising the first tranche of the Text Creation Partnership transcriptions of EEBO/ECCO books. The Jisc Historical Texts (JHT) Application Programming Interface (API) will be demonstrated and free logins will given for its Labs feature that enables direct algorithmic interrogation of its datasets. (As chair of its AB, the PI was instrumental in creating the JHT Labs service and the publication of its API to encourage such high-end exploitation of the resource.) The Hackathon will close with demonstrations of the most interesting work (as chosen by delegates) and small-value Job-Well-Done prizes for achievement at each level of expertise from beginner to expert. The Hackathon will be open to all attendees aged above ten years (11-16 year olds must bring an accompanying adult) with compulsory pre-registration and acceptance dependent upon online submission of a brief (<300 words) statement of interests; school-group registration will be encouraged. Free food and drink will be provided throughout the event, which will run non-stop from mid-day to mid-day.

3. Conference A three-day international conference on "Computational Methods for Literary-Historical Textual Studies" will be hosted by DMU to explore the latest methods and their discoveries and to address the challenges to computational stylistics arising from within and outside the field. (The most powerful

6

external challenges are commonly variants of Stanley Fish's humanistic rejection of stylistics *tout court*, which appears all the more penetrating--and cannot be ignored--when computers are doing the analysis.) A particular focus of this conference will be the ways that literary-historical scholarship will turn increasingly algorithmic in the future. Whereas the Hackathon encourages engagement at all levels and will be open to the public, this will be a traditional academic conference for dissemination of new discoveries in the field. A published version of selected conference papers and datasets will be offered as a special issue to the journals *Digital Humanities Quarterly* or *Digital Scholarship in the Humanities*.

References

Burrows, John "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship" *Literary and Linguistic Computing* 17 (2002): 267-87

_____ "Questions of Authorship: Attribution and Beyond" *Computers and the Humanities* 37 (2003): 5-32

_____ "All the Way Through: Testing for Authorship in Different Frequency Strata" *Literary and Linguistic Computing* 22 (2007): 27-47

Craig, Hugh "Shakespeare's Vocabulary: Myth and reality" *Shakespeare Quarterly* 62 (2011): 53-74

_____ The Date of *Sir Thomas More*" *Shakespeare Survey* 66 (2013): 38-54

Craig, Hugh and Arthur F. Kinney *Shakespeare, Computers, and the Mystery of Authorship* (Cambridge: Cambridge UP, 2009)

Egan, Gabriel "Shakespeare: Editions and Textual Studies" *The Year's Work in English Studies* edited by William Baker and Kenneth Womack (Oxford: Oxford University Press, 2001-)

_____ "Hearing or Seeing a play?: Evidence of Early Modern Theatrical Terminology *Ben Jonson Journal* 8 (2001): 327-47

_____ "Impalpable Hits: Indeterminacy in the Searching of Tagged Shakespearian Texts" Paper for the 33rd Annual Meeting of the Shakespeare Association of America, Bermuda, 17-19 March 2005

_____ *The Struggle for Shakespeare's Text: Twentieth-Century Editorial Theory and Practice* (Cambridge: Cambridge UP, 2010)

_____ "A History of Shakespearian Authorship Attribution" in Gary Taylor and Gabriel Egan (eds) *Shakespearian Authorship: A Companion to the New Oxford Shakespeare* (Oxford: Oxford UP, forthcoming 2016)

Egan, Gabriel, Steven Roth and Matthew Steggle "SHAXICAN: Perl Scripts Working on Shakespeare's Language" Internet. http://gabrielegan.com/SHAXICAN. 2002

Egan, Gabriel, Alejandro Ribeiro, Mark Eisen, and Santiago Segarra "Word Adjacency Networks and Shakespearian Authorship Attribution" *Shakespeare Quarterly* (accepted for summer 2016 issue)

Holdsworth, R. V. *Middleton and Shakespeare: The Case of Middleton's Hand in 'Timon of Athens'* Unpublished PhD thesis submitted to the University of Manchester (1982)

Hope, Jonathan *The Authorship of Shakespeare's Plays: A Socio-Linguistic Study* (Cambridge: Cambridge UP, 1994)

Irace, Kathleen O. *Reforming the 'Bad' Quartos: Performance and Provenance of Six Shakespearean First Editions* (Newark DE: U of Delaware P, 1994).

Jackson, Macdonald P. *Studies in Attribution: Middleton and Shakespeare* (Salzburg: Institut fur Anglistik und Amerikanistik, Universitat Salzburg, 1979)

_____ *Defining Shakespeare: 'Pericles' as Test Case* (Oxford: Oxford UP, 2003)

_____ *Determining the Shakespeare Canon: 'Arden of Faversham' and 'A Lover's Complaint'* (Oxford: Oxford UP, 2014)

Lake, David J. *The Canon of Thomas Middleton's plays: Internal Evidence for the Major Problems of Authorship* (Cambridge: Cambridge UP, 1979)

Rosso, Osvaldo A., Hugh Craig and Pablo Moscato "Shakespeare and other English Renaissance Authors as Characterized by Information Theory Complexity Quantifiers" *Physica A* 388 (2009): 916-26

Shannon, Claude & W. Weaver *The Mathematical Theory of Communication* (Urbana: U of Illinois P, 1949)

Shannon, Claude "Prediction & Entropy of Printed English" *Bell System Technical Journal* 30 (1951): 50-64

Taylor, Gary and Gabriel Egan (eds) *Shakespearian Authorship: A Companion to the New Oxford Shakespeare* (Oxford: Oxford University Press, forthcoming in 2016).

Vickers, Brian "Thomas Kyd: Secret Sharer" *Times Literary Supplement* 5481 (18 April 2008): 13-15

_____ "The Marriage of Philology and Informatics" *British Academy Review* 14 (2009): 41-44

_____ "Disintegrated: Did Thomas Middleton Really Adapt *Macbeth*?" *Times Literary Supplement* Number 5591 (28 May 2010): 14-15

_____ "Identifying Shakespeare's Additions to *The Spanish Tragedy* (1602): A (Newer) Approach" *Shakespeare* 8 (2012): 13-43

7

Dr Gabriel Egan

is Professor of Shakespeare Studies and Director of the Centre for Textual Studies at De Montfort University

researches on Shakespeare, computational methods, textual practice, theatre design/practice, bibliography, editorial theory

is currently working on computational approaches to authorship and General Editing the New Oxford Shakespeare (due 2016) having written *The Struggle for Shakespeare's Text* for Cambridge University Press (2010) and edited the collection *Electronic Publishing: Politics and Pragmatics*. His previous books are the *Edinburgh Critical Guide to Shakespeare* (2007), *Green Shakespeare* (2006), *Shakespeare and Marx* (2004, trans. 2007), and an edition of *The Witches of Lancashire* (2002)

co-edits the journal *Theatre Notebook* (ISSN 0040-5523) with Trevor R. Griffiths and Sarah McCleave and **co-edits the journal *Shakespeare*** (ISSN 1745-0918), with Deborah Cartmell, Lisa Hopkins, Brett D. Hirsch and Tom Rutter.

Points of contact

mail@gabrielegan.com

www.gabrielegan.com

School of Humanities
De Montfort University
Leicester LE1 9BH
United Kingdom
+44 116 257 7158

got good reviews for his last two books: *The Struggle for Shakespeare's Text* (2010) is "an incisively argued history of the set of ideas underlying Shakespeare's changing text" (Lukas Erne *Around the Globe* 48 (2011) p. 43); "a wonderfully thoughtful account of twentieth-century editing" (David Scott Kastan *Shakespeare Newsletter* 60 (2011) p. 25."The genius of" *Electronic Publishing: Politics and Pragmatics* was mentioned by a reviewer who thought "this inspirational volume" to be "well timed" (*N&Q* 258 (2013) p. 317). He also **got good reviews for *Green Shakespeare* (2006):** "this important book" (*Renaissance Quarterly* 60 (2007), p. 314); "intellectually challenging Green readings", "timely and necessary" (*Cahiers Élisabéthains* 70 (2007): pp. 81-2); "wonderfully illuminating", "wisely sensitive", "The analysis here ... is thoroughly compelling" (*Review of English Studies* 57 (2006), p. 820-2). There were also **good reviews for *Shakespeare and Marx* (2004):** "incisive and valuable readings of Shakespeare" and "well worth reading for Egan's innovative treatments of *Lear* and *Hamlet*" (*Times Literary Supplement* number 5312, 21 January 2005, p. 28).

was externally funded by the AHRC as Principal Investigator on the £416,293 two-year project "Shakespearean London Theatres (ShaLT)", a collaboration with the Victoria & Albert Museum to get tourists visiting and learning about the sites of late-16th and early-17th century theatres (September 2011 to September 2013). Previously got: £5,000 from JISC to build a printing press in Second Life (2009-10); paid fellowships at the Folger and Huntington libraries (2008-9); a British Academy grant (£3,600) to build a computer research model of the Globe playhouse to test the accuracy of its modern replica (2000-1); full British Academy studentships for his PhD (1994-7) and MA (1993-4).

serves as a technical reviewer for the peer review college of the Arts and Humanities Research Council (AHRC, 2008-), **was an assessor** for the JISC projects 'Digitization Project Phase Two' (2006-8) and 'Enriching Digital Resources' (2008-9). **He also chairs** the JISC Historical Texts (2011-) Advisory Board and the serves **on the steering group** for JISC's project Open Access Publishing in European Networks, UK (OAPEN UK) and **previously served** on the JISC E-Books Working Group (2005-7).

External Examining I examine the Open University's BA English course "AA306 Shakespeare: Text and Performance" (2009-present) and formerly examined De Montfort University's MA by Independent Study (2007-9).

De Montfort University, 1 June 2012 to Present: Professor of Shakespeare Studies In this post I serve also as Director of De Montfort's Centre for Textual Studies (CTS), a Digital Humanities Research Centre that has produced electronic editions of literary classics such as Geoffrey Chaucer's *The Canterbury Tales*, Thomas Malory's *Morte Darthur*, and Virginia Woolf's 'Time Passes'. I am currently developing the CTS's future research-project profile with a collection of proposed international collaborative projects to be supported by AHRC, British Academy, Leverhulme Trust, and National Endowment for the Humanities funding. I also serve as De Montfort's REF Unit of Assessment 29 Coordinator and Chair of the Ethics Committee.

Loughborough University, 1 August 2004 to 31 May 2012: Reader in Shakespeare Studies (Senior Lecturer until March 2008) In this post I served as my department's Research Coordinator from 2006 to 2008 (including the RAE submission) as well as Library Liaison and Environmental Representative. I have continued my research and publishing while also leading new undergraduate and postgraduate modules of my own devising, supervising five research students, and convening the MA Texts in Performance. I specialize in Shakespeare, bibliography, literary theory, and original staging of Renaissance drama and teach by lectures, personal tutorials, seminars, and workshops. For my students I create online resources of primary and secondary materials from the public domain and, where licensing arrangements permit, copyright-protected sources. Between 2010 and 2012 I developed a new, practical final-year module on hand-printing, using an Albion iron-press to print Shakespeare's sonnets on hand-made paper.

Shakespeare's Globe, July 2000 to July 2004: Lecturer I was hired to develop a new MA Shakespearean Studies with King's College London and teach its core course, and to devise and teach undergraduate Shakespeare courses. Of the first MA cohort of 13 students, two progressed to PhD studies at King's (both having secured full AHRB funding), and of the second cohort, two more registered for PhDs in Britain and two for PhDs in the USA. As I left this position in the summer of 2004, two more students from the course were awarded full AHRB grants to do their PhDs, confirming the degree's status as a world-class 'feeder' into research streams. I taught on the practices of Renaissance playing companies, theatrical and cultural history, and ran practical classes in handprinting using a replica wooden handpress

15 January 1996 to 30 June 2000 On a series of temporary contracts I taught at the University of Birmingham UK (15 January 1996 to 31 July 1997), East Carolina University USA (17 August 1998 to 7 January 1999), and De Montfort University UK (1 February 1999 to 30 June 2000)

Research Supervision Aside from overseeing all research within the Centre for Textual Studies at De Montfort University (of which I am Director), I am currently supervising 4 PhD students and in the past 5 years I have examined 5 PhD theses in the UK, the Netherlands, and the USA.

Dr Gabriel Egan's Publications

Authored Books

Egan, Gabriel. *Shakespeare and Ecocritical Theory*. The Arden Shakespeare. London: Bloomsbury. 2015.

* Egan, Gabriel. *The Struggle for Shakespeare's Text: Twentieth Century Editorial Theory and Practice*. Cambridge: Cambridge University Press, 2010.

Edited Books

* Taylor, Gary, Terri Bourus, John Jowett, and Gabriel Egan (eds) *The New Oxford Shakespeare* (Oxford: Oxford University Press, forthcoming in 2016)

* Taylor, Gary, and Gabriel Egan (eds) *Shakespearian Authorship: A Companion to the New Oxford Shakespeare* (Oxford: Oxford University Press, forthcoming in 2016)

* Egan, Gabriel (ed.). *Electronic Publishing: Politics and Pragmatics*. New Technologies in Medieval and Renaissance Studies. Toronto: Medieval and Renaissance Texts and Studies (MRTS) and ITER, 2011

Chapters in Books

*Egan, Gabriel "The Provenance of the Folio Texts" *The Cambridge Companion to Shakespeare's First Folio (1623)*. Ed. Emma Smith. Cambridge: Cambridge University Press. forthcoming in 2015.

Egan, Gabriel "Closure of the Theatres" *The Yearbook of English Studies 44*. Ed. Rory Loughnane, Andrew J. Power and Peter Sillitoe. London: Modern Humanities Research Association. 2014. 103-19.

* Egan, Gabriel. "The Presentist Threat to Editions of Shakespeare." *Shakespeare and the Urgency of Now: Criticism and Theory in the Twenty-First Century*. Ed. Cary DiPietro and Hugh Grady. New York: Palgrave, 2013. 38-59.

Egan, Gabriel. "Homeostasis in Shakespeare." *Posthumanist Shakespeares*. Ed. Stefan Herbrechter and Ivan Callus. London: Palgrave Macmillan, 2012. 77-94.

Egan, Gabriel. "*Sir Thomas More*." *The Facts on File Companion to Shakespeare*. Ed. William Baker and Kenneth Womack. New York: Infobase Learning, 2012. 1735-59.

Egan, Gabriel. "Gaia and the Great Chain of Being." *Ecocritical Shakespeare*. Ed. Lynne Bruckner and Daniel Brayton. Aldershot: Ashgate, 2011. 57-69.

* Egan, Gabriel. "Shakespeare, Idealism, and Universals: The Significance of Recent Work on the Mind". *The Return of Theory in Early Modern Literary and Cultural Studies: Tarrying with the Subjunctive*. Ed. Paul Cefalu and Bryan Reynolds. Basingstoke: Palgrave Macmillan, 2011. 278-295.

Journal Articles

* Egan, Gabriel, Alejandro Ribeiro, Mark Eisen, and Santiago Segarra "Attributing the Authorship of the *Henry VI* Plays by Word Adjacency" *Shakespeare Quarterly* (accepted and forthcoming in summer 2016)

* Egan, Gabriel "What is Not Collaborative about Early Modern Drama in Performance and Print?" *Shakespeare Survey* 67 (2014): 18-28.

* Egan, Gabriel. "Editorial Treatment of Press Variants: Q2 *Hamlet* as a Test Case" *Papers of the Bibliographical Society of America* 106 (2012): 311-55.

* Egan, Gabriel. "Precision, Consistency, and Completeness in Early Modern Playbook Manuscripts: The Evidence from *Thomas of Woodstock* and *John a Kent and John a Cumber*." *The Library* Seventh Series 12 (2011): 376-391.

* Egan, Gabriel. "Intention in the Editing of Shakespeare." *Style* 44 (2010): 378-90.

16/09/15

To whom it may concern

Rene Court
Coldhams Road
Cambridge
CB1 3EW

01223 214446

RE: Travelling Roadshow Performances on 'How Computers Work'

The Centre for Computing History, a public-facing educational charity, welcomes this opportunity to collaborate with De Montfort University on Professor Egan's Travelling Roadshow performances on 'How Computers Work'. This project aligns with the Centre's strategic aim to deliver inspirational learning opportunities to a wide range of audiences and will create an opportunity to expand our offer to education-based audiences and experiment with new ideas for our own workshops and future events. The project complements our current activities in education, such as workshops demonstrating various aspects of coding and electronics. It will complement these activities by providing the audience with an understanding of how a computer functions at a fundamental level, rather than limiting our focus to software or simple electronics. It may also allow us to engage an adult audience as well as a school-age one.

Prior to the Roadshow visiting Cambridge, the Centre for Computing History will advertise the event in digital and print media and will provide administrative support to the Prof Egan's team ahead of their arrival. We have costed this at £288. During the Roadshow in Cambridge, the Centre for Computing History will provide a suitable performance space for its two public performances including free use of rooms (£1,152), security and access control to buildings (£206), administration of event ticketing (£97), and front-of-house services immediately before, during, and after each show (£770).

Thus, if the proposed Roadshow comes to Cambridge to perform at our museum, the Centre for Computing History will provide in-kind contributions to a total of £2,513 in support of the two performances.

Yours faithfully

Jason Fitzpatrick

Curator

www.ComputingHistory.org.uk

Lucie Burgess MPhys (Oxon)
Associate Director for Digital Libraries

Clarendon Building, Broad Street, Oxford OX1 3BG
TEL: +44 (0)1865 277104   FAX: +44 (0)1865 277187
EMAIL: lucie.burgess@bodleian.ox.ac.uk

Bodleian Libraries
UNIVERSITY OF OXFORD

17 July 2015

Professor Gabriel Egan
School of Humanities
De Montfort University
The Gateway
Leicester, UK, LE1 9BH

Bodleian Libraries Statement of Support for the 'Travelling Roadshow on Computational Methods in Textual Studies'

Dear Professor Egan,

The Bodleian Libraries welcomes the opportunity to enter into a partnership with De Montfort University on the proposed 'Travelling Roadshow on Computational Methods in Textual Studies'. The Bodleian Libraries is committed to exploring the use of computational methods in the investigation of texts, and the assessment of tools making this possible, as is evident from the Libraries' support for the EEBO-TCP project. The Bodleian Libraries wishes to further encourage research, expertise, and entrepreneurship in the field of digital textual studies, and has created the Centre for Digital Scholarship within the new Weston Library to support these aims.

For this partnership, the Bodleian Libraries would commit to the following 'in kind' contributions to the project in support of a two-day training session to take place in the Weston Library, Bodleian Libraries:

Prior to the roadshow visiting Oxford, the Bodleian Libraries will advertise the event in digital and print media and will provide administrative support to manage delegate registrations and related logistics.

During the roadshow in Oxford, the Bodleian Libraries will provide suitable space and facilities for its events. Facilities provided free will be: use of the Centre for Digital Scholarship for a suitable number of delegates; use of another area for refreshments; IT provision and support up to the standard for office use and presentations, including digital projection; and provision of delegate packs (in digital format).

The proposed roadshow would bring tangible benefits to the Bodleian Libraries and hence we are happy to commit to these 'in kind' contributions to a value of £7,790.

Yours sincerely

Lucie Burgess
Associate Director for Digital Libraries

Itemized "in kind" contributions

All figures are exclusive of VAT

Centre for Digital Scholarship, Weston Library hire: 2 days x 7 hours @ £300/hour = £4,200

Refreshments area: 2 days x 3 hours @ £200/hour = £1,200

AV support: 2 days @ £400/day = £800

Designing print and digital advertising: 2 days @ £430/day = £860

Digital campaign: 1 day @ £430/day = £430

Print run: £300

Grand total = £7,790

Professor Steve Rothberg Pro Vice-Chancellor (Research)
Loughborough University Leicestershire LE11 3TU UK
Direct line: +44(0)1509 223440 Fax: +44 (0)1509 223900
E-mail: S.J.Rothberg@lboro.ac.uk

Loughborough
University

To whom it may concern

18th August 2015

RE: The Travelling Roadshow on Computational Methods in Textual Studies

Loughborough University welcomes this opportunity to collaborate with De Montfort University on Professor Egan's 'Travelling Roadshow on Computational Methods in Textual Studies'. This project aligns with Loughborough University's strategic aim to promote a vibrant research and scholarship culture and will create an ideal forum for interdisciplinary practice. The project complements current expertise in the field of textual studies, the study of publishing, and the evolving specialism of digital humanities. We look forward to pursuing new links and methods as an outcome of the Roadshow.

The proposed link, who will assist in demonstrations and lab-based hands on sessions, is Dr Clare Hutton, Senior Lecturer in English, an expert in the field of book history, who has recently completed a Leverhulme-funded project on 'The Textuality of Joyce's *Ulysses*' (proceeding to press with Oxford University Press under the title *Serial Encounters: Ulysses and the Little Review*). This work has involved the study of variant texts in an electronic environment using software packages such as Juxta and Tustep. Through training at the CTS, Dr Hutton is keen to develop and deepen her work in this area, and to gain a better understanding of the ways in which digital resources are transforming the possibilities of literary history (along the line of 'macroanalysis' and 'distant reading' as proposed by Matthew Jockers and Franco Moretti respectively).

For this partnership, Loughborough University will commit to the following 'in kind' contributions to the project in support of the Travelling Roadshow visiting the University:

* Prior to the roadshow visiting Loughborough, the University will release from their other duties a suitable candidate ('the Link') to spend a week in residence at the Centre for Textual Studies of De Montfort University in Leicester. The cost of covering this person's usual duties while these activities take place is £2,664.40.

* Prior to the roadshow visiting Loughborough, the University will advertise the event in digital and print media and will provide administrative support to manage delegate registrations and related logistics. We have costed this at £325.

2

* During the roadshow al Loughborough, the University will provide suitable space and facilities for its events, including IT provision and support up to the standard for office use and presentations, including digital projection; and provision of delegate packs (in digital format), together with use of a publicly licensed venue for the two performances and all related logistics (technical support, access control, ticketing, security and first aid). We have costed this at £3,055.00.

* During the Roadshow's visit to Loughborough, the University will relieve the Link from other duties in order to spend 20 hours assisting in the delivery of its content. The cost of covering this person's usual duties while these activities take place is £1,332.20.

* In the year after the roadshow visits Loughborough, the University will free the Link from other duties to spend a total of 40 hours on activities following up the work of the roadshow to embed expertise in computational approaches to textual studies within the university. We have costed this at £2,664.40.

The proposed roadshow would bring tangible benefits to Loughborough University and hence we are happy to commit to these 'in kind' contributions to a value of £10,041.00

Yours faithfully

Professor Steve Rothberg
Pro Vice-Chancellor (Research)

9 July 2015

University of Strathclyde Statement of Support for the 'Travelling Roadshow on Computational Methods in Textual Studies'

To whom it may concern

The University of Strathclyde welcomes the opportunity to enter into a partnership with De Montfort University on the proposed 'Travelling Roadshow on Computational Methods in Textual Studies'.

Strathclyde University is committed to interdisciplinarity and collaboration across all areas of research and teaching, and the development of digital methods in a Humanities context is a key part of strengthening research excellence, and providing our students with digital and numeracy skills. The Roadshow would complement and build-on our established digital methods base, which includes TextLab, a research-based project class taught by Computer and Information Sciences and English Studies. The opportunities offered by the Roadshow would allow us to extend digital skills across academic staff, in keeping with our values of boldness and innovation (http://www.strath.ac.uk/whystrathclyde/values/).

As a leading international technological university, we are committed to an active, and critically informed, engagement with technology. The Roadshow is an exemplary instance of this, and would strengthen existing links between Strathclyde and De Montfort staff, as well as building new ones.

For this partnership, the university would commit itself to the following activities and 'in kind' contributions:

* Prior to the roadshow visiting Glasgow, the University of Strathclyde will release from their other duties a suitable candidate ('the Link') to spend a week in residence at the Centre for Textual Studies of De Montfort University in Leicester. After the residency in Leicester and before the roadshow visits Glasgow, the Link will be freed from other duties for a total of 20 hours to prepare for the roadshow's visit. The cost of covering this person's usual duties while these activities take place is £600.

* Prior to the roadshow visiting Glasgow, the University of Strathclyde will advertise the event in digital and print media and will provide administrative support to manage delegate registrations and related logistics. We have costed this at £288.

* During the roadshow in Glasgow, the University of Strathclyde will provide suitable space and facilities for its events, including free: use of rooms (£880); IT provision and support (£469); security and access control to buildings (£380); administration of event registration; and provision of delegate packs (£600). During the roadshow, the Link will be freed from other duties to co-present the roadshow with Prof Egan (£600). We have costed this at £2,929 in total.

* In the year after the roadshow visits Glasgow, the University of Strathclyde will free the Link from other duties to spend a total of 40 hours on activities following up the work of the roadshow to embed expertise in computational approaches to textual studies within the university. We have costed this at £1,200.

The proposed roadshow would bring tangible benefits to the University of Strathclyde and hence we are happy to commit to these 'in kind' contributions to a value of £5,017.

Signed

Professor Richard J. Finlay

Head of School, Humanities

To whom it may concern

10 June 2015

RE: The Travelling Roadshow on Computational Methods in Textual Studies

Liverpool John Moores University (LJMU) welcomes this opportunity to collaborate with De Montfort University on Professor Egan's 'Travelling Roadshow on Computational Methods in Textual Studies'. This project aligns with LJMU's strategic aim to promote a vibrant research and scholarship culture and will create an ideal forum for interdisciplinary practice. The project complements current expertise in Shakespearean studies and digital humanities innovations in the Centre for Literature and Cultural History; we look forward to pursuing new links and methods as an outcome of the Roadshow. Specific projects for which the Roadshow is relevant include the Digital Archive of Working-Class Writing, which has access to around 230 memoirs contained in the Burnett Collection of Working-Class Autobiography (Rogers); individual textual research on records of early English drama (Graham); research into Asian-American literature (Adams); textual and discourse analysis of literary modernism (O Donghaile) and studies of popular magazine fiction from the 1910s (Cranfield).

Dr Elspeth Graham, Reader in Early Modern Studies, will act as the main link for the Travelling Roadshow. Dr Graham will assist in identifying and releasing a suitable candidate at an early career stage (PhD or postdoctoral) to spend a week in residence at the Centre for Textual Studies of De Montfort University in Leicester. This residency will support the career development of a researcher to gain skills in computational methods and in advocating for these innovations among researchers in wider disciplinary networks.

Prior to the roadshow visiting Liverpool, LJMU will advertise the event in digital and print media and will provide administrative support to manage delegate registrations and related logistics. Our corporate communications team will provide support for this on an in-kind basis. During the roadshow in Liverpool, LJMU will provide space and facilities for its events, including IT provision and support, administration of event registration and provision of delegate packs. A room-hire fee of £175 has been waived by the University and the in-kind contribution for promotion comes to £237 for time and consumables.

During the roadshow, the early career researcher will be freed from other duties to co-present the roadshow with Professor Egan. Similarly, in the year after the Roadshow, the researcher will spend a

ljmu.ac.uk SCHOOL OF HUMANITIES AND SOCIAL SCIENCES
John Foster Building, 98 Mount Pleasant, Liverpool, L3 5UZ

total of 40 hours on activities to embed expertise in computational approaches to textual studies within the University. To assist with this, their responsibilities will be temporarily shared among the teaching and administration team in the Centre for Literature and Cultural History. This is an indicator of the long term benefits the roadshow will bring to staff at LJMU and hence we are happy to commit to these in-kind contributions.

Yours faithfully

Dr Alice Ferrebe
Subject Leader for English
Liverpool John Moores University

1. Summary of Digital Outputs and Digital Technologies

The digital outputs of the concentrated research phase of the project will be a website containing reports (detailing the experiments and their discoveries), source texts (transcriptions of the Shakespeare quartos and Folio), and source code and binary executables for the software developed by the project. These will be freely available for anyone to download--under a GNU General Public License for software and a CC-BY licence for everything else--except where the rights of others prevent this; as last eventuality may occur if we cannot avoid employing proprietary software, which we will avoid wherever possible. The research phase of the project comprises experiments in computational stylistics applied to literary texts (the Shakespeare quartos and Folio) encoded as plain text (ASCII) and marked up in Standard Generalized Markup Language (SGML) and eXtensible Markup Language (XML) using software written by the project team in a variety of languages (primarily C, Perl, and Python) and running on standard office desktop computers using the Microsoft Windows, Apple OS, and Linux operating systems. Outputs from our software will be processed by standard off-the-shelf data processing and visualization software including the Open Source R software environment and, where necessary, proprietary software including MATLAB and SPSS software packages. These experiments and their outcomes are direct expressions of, and answers to, the project's research questions.

The leadership and dissemination phase of the project has no digital outputs other than reports and promotional videos added to the project website. The digital technologies of the leadership and dissemination phase of the project comprise large textual corpora and software packages that will be i) taught in the Travelling Roadshows, and ii) demonstrated and made available to delegates at the Literature Hackathon. The particular corpora and software for (i) will be determined by negotiation (via the Link) with the host institution for each of the Roadshows in order to meet its specific expressions of interests. The corpora for (i) will include the Text Creation Partnership (TCP), public-domain transcriptions of images from Early English Books Online (EEBO) and Eighteenth-Century Collections Online (ECCO) and Jisc Historical Texts. The software for (i) are likely to include an online XML editor and one or more online programming environment and text-processing environments, running in a browser rather than from locally installed executables in order to minimize the logistical burden on the host organization and the risk of local incompatibilities, and to enable attendees to follow up their training with homework experimentation without having to acquire and install software. The Jisc Historical Texts Labs service (JHT-Labs) will provide a baseline architecture available at all host institutions via a free trial for those that do not already subscribe, giving in one platform usable electronic texts (derived from TCP) and an online programming environment. The corpora and software for (ii) will be the same as (i) but with extensive locally installed software in the labs comprised Open Source and proprietary commercial XML editors, programming environments, statistical packages, and visualization tools.

2. Technical Methodology

2a: Standards and Formats The base transcriptions of the Shakespeare quartos and Folio editions will be in Text Encoding Initiative (TEI) XML format--as used by our sources for these transcription--which will be the native format used for those experiments for which the element-structure of TEI-XML (*scene*, *speech*, *line*, and so on) is appropriate; for other experiments we will transform the TEI-XML to different formats as necessary, including plain ASCII. We choose TEI-XML because it is widely understood by humans and software systems, is well-defined, and saves us labour when we need to automate the division of our texts for various purposes such as separating out all the speeches of one speaker or of one kind (such as all prose from all verse). The typical TEI-XML file size for one Shakespeare play is 25MB and we will use around 40 such files (one per play).

2b: Hardware and Software The hardware for our experiments will be standard desktop computers running the operating systems Microsoft Windows, Apple OS, and Ubuntu Linux. Different applications and programming languages run best on different operating systems, but all run on one or more of these three. The software for our experiments will be written by the PI and PDRA as part

1

of the project in a range of languages including C, Python (for its Natural Language Toolkit), Perl, Prolog, C, Lisp, R, and MATLAB. We will likely use software written for other projects such as the Intelligent Archive from the University of Newcastle, Australia, and Lancaster University's VARD-2 system, both written in the language Java and useful for their normalization of early modern spelling variants. When we employ others' software we will always prefer Open Source projects so we can check the source code to make sure it does what its creators' think it does.

2c: Data Acquisition, Processing, Analysis and Use The data acquisition phase for this project is the acquiring of the various transcriptions of the Shakespeare quartos and Folio. We will download freely available transcriptions--including public domain texts from EEBO-TCP Phase One--and receive from the Internet Shakespeare Editions (ISE) project its high-quality, hand-crafted transcriptions of the Shakespeare quartos and Folio, as already agreed with ISE. The ISE texts are currently in Standard Generalized Markup Language (SGML) but are likely to have been transformed to XML by the time the project commences; we can comfortably handle either format. Using the variant-detection capabilities of the software Juxta and TUSTEP/TXSTEP the transcriptions for each early edition will be compared one-with-another in order to satisfy ourselves that any differences between them--arising, for example, from stop-press correction in the exemplars used and from editorial interpretation of illegible characters--will not significantly affect our experiments; the transcriptions will be hand-emended if necessary and the changes documented.

The processing, analysis and use of these data will use the software that we write for the computational stylistics experiments described in the Case for Support. We will start with our own implementation of standard algorithms for the following tests:

1) Measurement of the Shannon Entropy and Jensen-Shannon Divergence of texts to produce Statistical Complexity Measurements of the kinds recently demonstrated as authorship-style markers (López-Ruiz, Mancini & Calbet 1995; Martín, Plastino & Rosso 2003; Rosso, Craig & Moscatoa 2009)
2) Creation of Word Adjacency Networks using Markov chains to store the proximity values of 100+ function words found within a text (Egan et al. 2016)
3) Nearest Shrunken Centroid (Jockers & Witten 2010)
4) Random Forests (Breimen 2001)
5) Burrows & Craig's Delta, Zeta, and Iota (Burrows 2002, 2003, 2007; Craig & Kinney 2009)

For several of these tests there exist software implementations created by their inventors, but to ensure that the tests are doing what their creators believe them to be doing we will write our own implementations from the published algorithms in order to provide a check on investigator error.

All phases of the experiments from the writing of software to the collation and publishing of results will be documented in a log authored by the project's PDRA and check by the PI that will appear weekly on the project website. (This relatively frequent documentation cycle is essential for software development in particular, where many small decisions made in rapid succession may easily have large effects on the results.) Over the period of the project the weight of responsibility for software development and experimental design will steadily shift from the PI the PDRA as the latter acquires the necessary skills and expertise. Progress in software development, experiment design, and experiment execution will be reported fortnightly to the project Advisory Board with whom a monthly conference call will be held to collate feedback, discuss practical and theoretical problems arising from the work, and plan adjustments to the experiments--and if necessary adjustments to the Workplan--in the light of what we learn.

The desktop computers used for these experiments will contain the base transcriptions, the source code, and project reports and notes on high-speed networked drives that are mirrored by the University's IT department to RAID drives held centrally and backed up daily to offsite storage.

Section 3: Technical Support and Relevant Experience The PI (Egan) will be the single most important repository of the project's technical expertise. Egan has 35 years practical experience in computer programming, including 10 years of industry experience before entering Higher Education. His expertise is drawn upon by Jisc, for whom he has served as technical advisor on the 'E-Books Working Group' (2005-7), 'Digitization Project Phase Two' (2006-8) and 'Enriching Digital Resources' (2008-9), and 'Open Access Publishing in European Networks, UK (OAPEN UK)' (2012-present), and he chairs the Advisory Board of 'Jisc Historical Texts' (2011-present). His technical expertise is drawn upon in evaluating proposals for papers at the annual conference of the Alliance for Digital Humanities Organizations (ADHO) and submissions to the journals *Literary and Linguistic Computing* (recently renamed *Digital Scholarship in the Humanities*) and the *Journal of the Text Encoding Initiative* and in his evaluation of funding bids for the AHRC, for whom he serves as a Technical Assessor in the Peer Review College. The selection process for appointing the PDRA will seek someone highly technically competent with an understanding of and interest in the linguistic nature of literary (preferably dramatic) texts.

The project's next line of defence on technical matters will be its Advisory Board, whose members have been chosen to create the broadest possible base of technical ability. Of those Advisory Board members: MacDonald P. Jackson has been publishing his research on stylistic approaches to authorship attribution since before computers were used to implement them (over four decades); John Burrows is the inventor of several of the computational stylistics techniques used in this project; Hugh Craig has refined Burrows's techniques and collaborated in their implementation as published computer source code; Brett D. Hirsch is a highly skilled early-career investigator who has shown how techniques from outside literary studies (gene-classification research and machine learning more widely) may be used for authorship attribution, and Doug Duhaime is a technical developer (with a PhD in literary studies) working for the ProQuest corporation that provides universities with the large corpora Literature Online (LION) and Early English Books Online (EEBO). A fifth member of the Advisory Board will be elected by its members after a discussion to identify remaining technical areas in which its strengths might usefully be increased. Across its members the project team already embodies more technical expertise than is available from the IT department of the project's home institution, and indeed the PI's Centre for Textual Studies is a source of technical expertise upon which the home institution routinely draws.

The project plans have been discussed with members of the Centre for Literary and Linguistic Computing (CLLC) at the University of Newcastle, Australia, the Centre for Textual Studies and Digital Humanities (CTSDH) at Loyola University, Chicago, and the Department of Electrical & Systems Engineering at the University of Pennsylvania. Their most useful advice has been collated and the present plans were shaped accordingly. No further training or expertise needs are anticipated.

Section 4: Preservation, Sustainability and Use The mere preservation of the materials created by the project will be ensured by a robust daily backup regime including off-site storage and the sustainability and reuse of them will be ensured by the scrupulous documentation of what we do and what we create and the dissemination in digital form of this documentation and our source code, lab notes, raw datasets, and published results, made available to all interested parties by online self-publication and non-embargoed deposit in the host institution's Institutional Repository. All the digital outputs of the project are worth preserving and disseminating and none will be excluded.

The project will be documented by a website created by the Centre for Textual Studies (CTS) at the host institution, De Montfort University, which undertakes to maintain this website and ensure its presence on the Worldwide Web for a minimum of 5 years after project funding ends. (The CTS specializes in the maintenance of legacy websites and has developed the technology that enables its continued hosting of large digital resources created at De Montfort in the 1990s and early 2000s with funding from the AHRB, as it was then, the AHRC, and the British Academy.) The website will hold all materials created by the project and make them available under a Creative Commons CC-

3

BY licence (for textual and visual materials) and a GNU Public License (GPL) for our source code and, where it may be convenient for other investigators, compiled binaries of our source code.

The project's methods have been designed for sustainability, reuse, and (we hope) replication by other investigators. The best means for ensuring that our results are reliable contributions to knowledge is encouraging other investigators to pore over everything we have done looking for logical flaws, errors in coding, and unjustified assumptions. This scrutiny is our prime motivation for making everything available under the least restrictive CC-BY and GPL agreements. We hope that others will take our code and run it for themselves, either on our datasets or their own, in order to find out how we might have done better or--in the worse case scenario--what we simply got wrong. We intend also that others wishing to enter this new field will find our online resources a convenient base from which to start their own investigations.

Our choices of data formats and programming languages are fundamentally shaped by these considerations of scrutiny and reuse. They are the reason that we prefer our base texts to be in TEI-XML (a standard ratified by the International Standards Organization, ISO) and why we will always use non-proprietary technologies where possible. In our programming, we will where the choice arises always prefer long-standing languages that many people understand (for example C and Perl) over newer ones that might offer attractive features but for which the known 'user-base' is likely to be smaller (for example Ruby).

References

Breimen, Leo "Random Forests" *Machine Learning* 45 (2001): 5-35

Burrows, John "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship*" Literary and Linguistic Computing* 17 (2002): 267-87

_____ "Questions of Authorship: Attribution and Beyond" *Computers and the Humanities* 37 (2003): 5-32

_____ "All the Way Through: Testing for Authorship in Different Frequency Strata" *Literary and Linguistic* Computing 22 (2007): 27-47

Craig, Hugh and Arthur F. Kinney *Shakespeare, Computers, and the Mystery of Authorship* (Cambridge: Cambridge University Press, 2009)

Egan, Gabriel, Alejandro Ribeiro, Mark Eisen, and Santiago Segarra "Word Adjacency Networks and Shakespearian Authorship Attribution" *Shakespeare Quarterly* (accepted for summer 2016 issue)

Jockers, Matthew L. and Daniela M. Witten "A Comparative Study of Machine Learning Methods for Authorship Attribution" *Literary and Linguistic Computing* 25 (2010): 215-24

López-Ruiz, R. L., H.L. Mancini, and X. Calbet "A Statistical Measure of Complexity" *Physics Letter A* 209 (1995): 321-26.

Martín, M. T., A. Plastino, and O. A. Rosso. "Statistical Complexity and Disequilibrium" *Physics Letters A* 311 (2003): 126-32

Rosso, O. A., H. Craig, and P. Moscatoa " Shakespeare and other English Renaissance Authors as Characterized by Information Theory Complexity Quantifiers" *Physica A* 388 (2009): 916-26

Pathways to Impact

The 12-month concentrated research phase of the proposed project will generate impact by giving critics and editors of Shakespeare's works a better understanding of the early editions upon which modern ones are based, leading to better knowledge of what he wrote, better editions of Shakespeare, and thus an enhanced intellectual experience for his many readers, students, teachers and performers across the world. The knowledge imparted will take the form of a book, a project website, and an academic conference. The website will house all the digital outputs of the project, including reports (detailing the experiments and their discoveries), source texts (transcriptions of the Shakespeare quartos and Folio), and source code and binary executables for the software developed by the project. These will be freely available for anyone to download, using a GNU General Public License for the software and a CC-BY licence for everything else.

The 6-month leadership and dissemination phase that follows the concentrated research phase (thus months 13-18 of the project) will generate impact via a series of training events for academics and a series of live-performance public-engagement events including interactive demonstrations that will transfer to their audiences practical knowledge about the use of computational methods for Textual Studies. The impact events of the leadership phase are:

> 1) Training in computational methods for academics, delivered as part of a Travelling Roadshow visiting Glasgow, Liverpool, Loughborough, Oxford, and a site in the South West (to be determined).

> 2) Two theatrical performances for the public delivered as part of the Travelling Roadshow visit to each of the above five towns (thus 10 performances) plus an additional pair of performances at the Centre for Computing History, a public museum in Cambridge.

> 3) A Literature Hackathon in Leicester open to all comers.

The enormous potential for the application of computational techniques to academic Textual Studies--that is, English, Drama, History and related text-based disciplines--is largely untapped because most textual scholars lack the foundational knowledge needed to make a start. The PI straddles the fields of literary-dramatic studies and computation and has designed this project to transfer computational knowledge and techniques into the largely new context of Textual Studies. The impact will thus mainly comprise new links between hitherto under-connected fields, evidenced in new projects that span these fields and new multi-disciplinary techniques applied to existing projects.

The training for (1) will be hands-on and based upon a bespoke curriculum developed in coordination with a person (called 'the Link') from each host institution in order to respond closely to local needs and interests. The topics covered will range from the simplest kinds of search-and-replace operations to reformat and restructure writing--one's own and one's primary texts--through the various kinds of counting and classifying of linguistic features, to the most sophisticated kinds of automated variants detection, Natural Language Processing, data visualizations, and authorship attribution by stylistics.

There will be one Link from each of the host institutions--the University of Strathclyde, Liverpool John Moores University, Loughborough University, the Bodleian Library, and another to be decided--and each host is committed (see Letters of Support) to their Link's own dissemination plan for bringing back to their institution the computational knowledge gained and skills developed on this project, including the Link's one-week, training-intensive secondment to the PI's Centre for Textual Studies in Leicester. Thus the Links provide a further pathway for the project to generate

impact within the hosts. For textual academics these activities achieve impact by enabling them to undertake new kinds of analysis not possible without computers.

For the wider public these activities achieve impact by raising awareness of the fundamentals of computer operation that make them amenable to text processing and analysis. A project website will support the dissemination of the research-phase discoveries and the leadership-phase training and live-performance events, providing a record of what has been done, tasters of the events before they happen, and records of how they went and what they achieved.

As a complex collaboration between five universities (DMU and four hosts of the Roadshows} and a research library (the Bodleian as fifth Roadshow host) and a museum (the Centre for Computing History) the project requires careful management to maximize impact. This is achieved by:

> 1) Extensive collaborative design and planning of the impactful activities with the project partners, which has already been undertaken and is embodied in specific and detailed mutual commitments laid out in the Letters of Support.

> 2) The PI, who is uniquely qualified to lead this endeavour, devoting 50% of his time to the project in the 6-month leadership and dissemination phase in order to fine-tune the activities as they occur. The PI developed his expertise for this work as PI on an £416,000 AHRC-funded Knowledge Transfer Fellowship in partnership with the Victoria & Albert Museum in 2011-13, and developed his theatrical-pedagogical expertise while Globe Education Lecturer at the Shakespeare's Globe theatre (London) in 2000-4.

The costliest resources expended by the project's impact activities will be the PI's time (paid for from the grant) and the time of the five Links and their institutions' administrators devoted to supporting the work (both paid for by in-kind contributions from their institutions). The next largest resources are the time and expenses (travel and subsistence) of the actors (the PI's students and/or former students) delivering the 12 live performances, and the expenses of running the Literature Hackathon. See Justification of Resources for details of all these.

For the in-house academic training the promotional targets are modest: news of the opportunity must reach academics in text-based subject disciplines. For the live-performance public-engagement events a much greater promotional effort is needed and for this the Publicity Office of each host is committed to a local media campaign including targeted press releases for TV, radio, newspapers, and news websites, supported by traditional printed artefacts (posters, banners, targetted flyers) and buzz-spreading via social media. The DMU Publicity Office in coordination with its Square Mile public-engagement arm <http://www.dmu.ac.uk/about-dmu/dmu-square-mile> will handle targeting of the audiences for the Literary Hackathon. DMU already has extensive partnerships with local schools, including the PI's own Computing Club, and these will form the base from which the Hackathon promotion will begin.

The impact targets of the project are extensive and ambitious, and performance will be measured by a range of feedback techniques. Each of the five academic host institutions is committed to allocating 40 hours of each Link's time in the year after funding ends to following up the Roadshow activities, including reporting to the PI on the progress of each Link's impact work. (Examples of success would include new projects undertaken using the skills imparted during the Roadshow and new computational methods being applied to existing textual projects.) For the public performances and the Literature Hackathon, the PI's expertise gained as AHRC Knowledge Transfer Fellow in 2011-13 will be exploited in the creation of print-based and online feedback mechanisms--including head-counts, pre/post-performances questionnaires, and invited video-log uploads--for capturing how these events have changed their audience's thinking and enabled them to undertake activities not previously open to them.

Workplan

Legend: PDRA = Postdoctoral Research Associate; FTE = Full Time Equivalent; PI = Principal Investigator; AB=Advisory Board

Project month ► / Activity ▼	-3 to 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Research Phase: PDRA at 100% FTE, PI at 100% FTE		▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓						
Leadership Phase: PI alone at 50% FTE														▓	▓	▓	▓	▓	▓
Advertising, interviewing & selection of PDRA	▓																		
AB completed by election & invitation	▓																		
Preparation of lab & office space for PDRA	▓																		
PDRA devotes time to reading (80% at start declining to 20% at end)		▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓						
Computational experiments performed & reported monthly to AB		▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓						
Quarterly AB meeting				▓			▓			▓			▓			▓			▓
PDRA career development meetings			▓		▓		▓		▓		▓		▓						
Milestone #1: PDRA performs first self-devised experiment				█															
PI & PDRA co-author one chapter for their book					▓			▓				▓	▓						
Milestone #2: PDRA devises half the experiments							█												
PI authors Briefings for AB & website publication							▓					▓							▓
Milestone #3: Typescript of PI & PDRA's book goes to publisher														█					
PI devises Travelling Roadshow events															▓				
Roadshow visits Venue#1																▓			
Roadshow visits Oxford																	▓		
Roadshow visits Liverpool																	▓		
Roadshow visits Glasgow																		▓	
Roadshow visits Venue#5																		▓	
Literature hackathon at DMU																			▓
International conference in Leicester																			▓
Milestone #4: Project website finalized with posting of final report																			█