

4. Cultures of Knowledge in transition: Early Modern Letters Online as an experiment in collaboration, 2009-2018

Howard Hotson, with Arno Bosse, James Brown, Miranda Lewis, and Elizabeth Williamson

1. Preamble

Cultures of Knowledge (CofK) is not a typical research project.¹ It has not been devoted from the outset to testing a historical thesis or producing a set of traditional research publications. For most of its history, its main objective has been to pilot a collaboratively populated, digital union catalogue of early modern learned correspondence; but its intentions are not exhausted by resource creation or IT systems development. Instead, CofK's deeper purpose has been to conduct a social and cultural experiment investigating the conditions in which individual scholars, projects, and institutions are willing to pool their data. Its research objective has been to discover the scholarly standards, technical systems, social relations, and cultural norms required for a community of scholars with many needs and problems in common to collaborate in unprecedented ways. The title of this essay, 'Cultures of Knowledge in Transition', is therefore intended to refer to more than the evolution of a single research project: instead, it views the project's history as a window on the broader transition taking place as research communities develop means of tackling problems far too large, not only for the lone researcher or individual research project, but for single institutions, disciplines, or even national academic communities to solve on their own.

Amongst the aspects of traditional humanistic research culture which collaborative projects currently are disrupting is the convention of sole authorship. This chapter narrates the activities and accomplishments of a large team led by CofK's project managers, James Brown, Elizabeth Williamson, and Arno Bosse, and EMLO's longstanding Digital Editor, Miranda Lewis. Hence, although this chapter has been written by a single hand, authorship is shared and the first-person plural is employed throughout.

2. Defining the project: February – December 2008

The genesis of Cultures of Knowledge can be precisely dated. In February 2008, the Andrew W. Mellon Foundation announced its intention to send a delegation to Oxford in search of 'humanistic research projects capable of transforming disciplines'. The soundings undertaken by the recently arrived Professor of the History of Science, Pietro Corsi, had revealed a concentration of expertise on the intellectual history of the seventeenth century deeply rooted in the university's history. The Oxford Philosophy Club, which met in Wadham

¹ 'Cultures of Knowledge' is a collaborative research project at the University of Oxford funded by the Andrew W Mellon Foundation in four phases between 2009 and 2018. During Phase I (Jan. 2009-Dec. 2012), the project was subtitled 'An Intellectual Geography of the Seventeenth-Century Republic of Letters' and documented its activities at <https://cofk.history.ox.ac.uk/index.html>. During Phases II (Jan. 2013-Apr. 2015), III (Apr. 2015-Apr. 2017), and IV (Apr. 2017-June 2019), the project was subtitled 'Networking the Republic of Letters, 1550-1750' and created a new website at <http://www.culturesofknowledge.org> (last accessed 5 June 2019). The authors are greatly indebted to Simon Burrows and Glenn Roe for their help in boiling this account of the many phases of this project down to its current size.

College in the 1650s and 60s, was one of the most fertile seedbeds from which the Royal Society sprang. Over time, the papers of many Oxford-educated Fellows of the Royal Society gravitated back to the Bodleian Library and Ashmolean Museum, exercising a magnetic attraction for later scholars. With the 350th anniversary of the Royal Society looming in 2010, this concentration of expertise and material provided a solid foundation for an ambitious collaborative undertaking.

Catalogues of learned correspondence relating to the early Royal Society were the centrepiece of the project from the outset; but many questions remained. Should the inventory also include documentation of other forms of learned exchange, such as libraries and their catalogues? Should it expand beyond Oxford men of science to embrace the republic of letters more generally? Should it be devoted primarily to hard-copy editorial work, or experiment with new digital media? A key role in answering these and related questions was played by a dialogue with the Mellon Foundation, which must in turn be situated within three crucial developments of those years.

The first of these developments was political. In 2004, only four years before the project was conceived, the European Union had enlarged dramatically to embrace a huge swathe of former Warsaw Pact countries. The prospect of helping to mend the divide between ‘eastern Europe’ and ‘the West’ was particularly irresistible to those studying displaced central Europeans like Samuel Hartlib (c. 1600-1662) and Jan Amos Comenius (1592-1670). It also appealed to Mellon’s senior vice-president, Harriet Zuckermann, a leading sociologist of science whose late husband, Robert K. Merton, had helped inspire work on Hartlib’s circle.² With Mellon’s encouragement, the mid-century ‘universal reformers’ were grafted onto the core projects on the early Royal Society; and the Department of Comenius Studies in the Institute of Philosophy of the Academy of Sciences of the Czech Republic in Prague became our first international partner.

The second seismic development was technological: the explosive growth of the internet. The first experiments in digitizing early modern correspondences – The Hartlib Papers Project, for instance, and the Athanasius Kircher Correspondence Project – had paid the price of digital pioneers. The operating systems and storage technologies on which they relied were rapidly rendered obsolete, raising huge questions about the sustainability of digital publication. Prudence demanded a compromise: texts edited in the project, which took up the greater part of CofK’s Phase I budget, would be published in hard copy, while the digital experiment would be confined to catalogue metadata. For the whole of Phase I, it remained unclear whether the core of the project was new media or old, until the project was transferred to the Scholarly Communications Division in Phase II, which is devoted primarily to the application of information technology to developing transformative scholarly tools, infrastructures, and networks. This Division’s ‘step-by-step, phased funding approach for planning, proof-of-concept, implementation, and integration into the broader system of scholarly communication’ has shaped the pace and direction of the project ever since.³

The third global development was economic: the financial crisis precipitated by the collapse of Lehman Brothers in September 2008. The negotiation of our final budget with Mellon took place during the period of radical exchange rate fluctuation which followed. By sheer good fortune the dollar was at a record low against the pound when the budget was

² Robert K. Merton, *Science, technology and society in seventeenth-century England* (New York and London, 1970); Charles Webster, *The Great instauration: science, medicine and reform 1626-1660* (Oxford, 2002); I. Bernard Cohen, ed., *Puritanism and the rise of modern science: the Merton Thesis*, (New Brunswick and London, 1990).

³ <https://mellon.org/programs/scholarly-communications/> (last accessed 7 June 2019).

finalised (in dollars) and a record high when the grant arrived. The resulting budgetary surplus, fully one third of the original value of the grant in Sterling, provided a powerful shock absorber against the inevitable contingencies of a speculative project. In the course of Phase I, it became apparent that we had underestimated the administrative workload needed to manage such a large and complicated project, the editorial workload needed to standardize tens of thousands of letter records created independently by multiple scholars without shared standards, and the technical difficulty of ingesting data assembled with all kinds of (sometimes obsolete) software. Without the flexibility provided by the exchange-rate windfall, it is highly unlikely that the transformative potential of the project would have become sufficiently clear by 2012 to open the way to further rounds of funding.

3. Generating fresh data in house: Phase I, 2009-2012

3.1 Digitizing the Bodleian's main catalogue of early modern manuscript correspondence

The most basic scholarly objective of the project was an inventory of learned correspondence from the long seventeenth century. A natural point of departure was the digitization of the most comprehensive existing catalogue of early modern manuscript correspondence in the Bodleian Library. The Index of Literary Correspondence was a card catalogue of some 487 volumes of manuscript letters mostly from the seventeenth and eighteenth centuries. Located in a file cabinet, this rich resource had remained little known outside Oxford. An online digital version would allow scholars anywhere to search it not only alphabetically by sender, but by all other record fields, including the detailed abstracts of each letter laboriously recorded on the cards.

The scanning and keying of all 65,520 cards, overseen by the Bodleian's Head of Digital Collections, Mike Popham, was completed on budget and roughly on time in September 2009. What proved wildly optimistic was the plan to cleanse and standardise tens of thousands of records in a mere nine months between October 2009 and May 2010. As Miranda Lewis has shown,⁴ the card catalogue was compiled over several decades by three volunteers, each working independently and without prior agreement on cataloguing standards. Confronted by these difficulties, the project began incrementally to develop far more sophisticated infrastructure than originally planned and costed.

As a result, the original aims of this aspect of the project were exceeded in three respects. First, instead of merely processing the 22,592 letters written before 1712, all 48,691 records were scanned and keyed in Year 1 and uploaded in Year 2. Second, the basic metadata for each record was not merely de-duplicated and standardized, it was also enhanced with thousands of separate person records (recording occupations, familial and professional relationships, editorial notes, and links to related resources) and place records (recording varieties in nomenclature, geographical coordinates, links to GeoNames, and location within a seven-level hierarchy). Third, as described in section 4 below, all of this required extensive upgrades to the digital infrastructure previously envisaged.

⁴ Lewis, 'Ghosts in the Machine: (Re)Constructing the Bodleian's Index of Literary Correspondence, 1927-1963': CofK blog, 26 Apr. 2013. http://www.culturesofknowledge.org/?page_id=6 (last accessed 7 June 2019). All following references to the CofK blog can be found at the same page.

3.2 Six new catalogues

The project's lead scholars simultaneously began compiling inventories of five key correspondences, to which was later added a sixth. The correspondence of John Aubrey (1626-97) was a flagship project and also something of an outlier, since this inventory alone was created from scratch. One of the foremost gatherers and disseminators of scholarly and biographical information of his day, Aubrey was author of *Brief Lives*, a pioneer of archaeology and antiquarianism, and a member of the Royal Society. Despite their importance, less than one in ten of Aubrey's letters had ever been edited and published, and frequent use endangered the manuscripts. In the course of a doctoral dissertation supervised by Rhodri Lewis and William Poole, Kelsey Jackson Williams compiled a calendar of 1,073 letters, which were published on EMLO with 526 high-resolution scans.⁵

For all other calendars published during Phase I, Mellon funding helped to bring longstanding projects closer to conclusion. Philip Beeley brought with him to Oxford in 2008 the task of editing the correspondence of Oxford's Savilian professor of geometry, John Wallis (1616-1703). Despite the death of his 83-year-old co-editor, Christoph Scriba, in 2013, Philip published their calendar of 2,002 records in August 2011 and two planned volumes on schedule in 2012 and 2014.⁶

The original application also called for a calendar of the correspondence of Martin Lister (1639-1712), a pioneering natural historian and vice-president of the Royal Society. The discovery that Dr Anna Marie Roos was already working on an edition of Lister's correspondence allowed us to enhance this project further. Her calendar formed the basis of a three-volume hard-copy edition of Lister's correspondence, the first volume of which was awarded the John Thackray Medal of the Society for the History of Natural History in 2017.⁷

In 1996, a pioneering effort in digital editing, the Hartlib Papers Project in Sheffield, had published transcriptions of the papers of the Anglo-German intelligencer, Samuel Hartlib, on two CD-ROMs. Navigation of this rich trove of material was hindered, however, by the lack of a proper inventory. Assisted by a small grant in 2000, Howard Hotson began to create a database of the correspondence in Hartlib's archive which by 2009 had grown to 4,263 entries.⁸ Whilst checking this inventory against the original manuscripts, CofK's first post-doc, Leigh Penman, discovered over 450 hitherto unlisted Hartlib letters, mainly in the Bodleian, British Library, and printed dedications. Meanwhile, Mark Greengrass had obtained funding from the Mercers' Company to republish the images and transcriptions of the Hartlib Papers online on open access in the Humanities Research Institute of the University of

⁵ Kelsey Jackson Williams, *The Antiquary: John Aubrey's historical scholarship* (Oxford, 2016). CofK also financially supported the publication of William Poole, *John Aubrey and the advancement of learning* (Oxford, 2010). Further outputs listed at <https://cofk.history.ox.ac.uk/projects/john-aubrey/> (last accessed 7 June 2019).

⁶ *The Correspondence of John Wallis (1616-1703)*, eds. Philip Beeley and Christoph J. Scriba, vols. III (October 1668-1671) and IV (1672-April 1675) (Oxford, 2012-2014). Additional information at <https://cofk.history.ox.ac.uk/projects/john-wallis/> (last accessed 7 June 2019).

⁷ *The Correspondence of Dr. Martin Lister (1639-1712)*, ed. Anna Marie Roos, vol. 1: 1662-1677 (Leiden, 2015). Her half-time research appointment left time for an exceptionally rich range of outputs, dissemination activities, and spin-off grant applications, most notably the intellectual biography, *Web of Nature: Martin Lister (1639-1712), the First Arachnologist* (Leiden: Brill, 2011). See further <https://cofk.history.ox.ac.uk/projects/martin-lister/> (last accessed 7 June 2019).

⁸ The grant was from the Foundation for Intellectual History.

Sheffield.⁹ Michael Pidd and Jamie McLaughlin then assisted the CofK team in linking each letter record on EMLO to the respective transcription in Sheffield. As well as finally providing an inventory of Hartlib's letters, this partnership demonstrated how EMLO could become a meeting place for interlinked full-text correspondences hosted on project servers throughout the world.

The difficulty of reassembling the correspondence of Jan Amos Comenius was disproportionate to its relatively small size. After an adult life spent mostly in exile, Comenius' correspondence ended up scattered in repositories in at least thirty-two European cities. Comenius scholars at the Czech Academy of Sciences had assembled a database of over 500 extant letters printed in over thirty separate publications.¹⁰ Under the auspices of CofK, Kateřina Horníčková and Iva Leškova, working under the supervision of Vladimír Urbánek, checked and upgraded metadata, translated abstracts, acquired high-resolution scans and sought permission to reproduce them on EMLO. The resulting catalogue (currently consisting of 566 letter records) is amongst the richest in EMLO, including abstracts, people and places mentioned, incipits and explicits, and over 400 digital facsimiles. This upgraded catalogue forms the basis for a new, critical, and fully annotated edition of Comenius's complete correspondence, being published as part of his *Opera Omnia* by the Czech Academy of Sciences.¹¹

Finally, the exchange-rate surplus allowed us to incorporate a sixth correspondence: the letters of Edmund Lhwyd (1660–1709), second keeper of Oxford's Ashmolean Museum and a pioneer of Welsh philology and archaeology. Brynley Roberts had spent his career at the University of Wales's Centre for Advanced Welsh and Celtic Studies assembling transcriptions of Lhwyd's letters. With the financial and infrastructural support of CofK, Editorial Assistant Helen Watt, working under the supervision of Dafydd Johnston and Richard Sharpe, completed a calendar of 2,128 letters to or from Lhwyd and, with Brynley Roberts, transcriptions of the entire corpus in 2015.¹²

4. System building: EMLO as in-house pilot project

4.1 Catalogue Object Model

The basic precondition for both scholarly and technical work was a robust model for correspondence metadata. The key to the development of the EMLO data model was our original decision to distinguish between core and supplementary metadata.

According to this distinction, core data fields are those implicit in a straightforward description of what a letter is. In essence, a letter is a written communication from one or more persons in one place at a specified time to another person or persons at another place. In order to distinguish one letter from another, it is ordinarily sufficient to record only six data points: the sender, place of sending, date of sending, intended recipient, place of receipt, and location of the letter itself or record of it. Founding a union catalogue of correspondence on

⁹ <https://www.dhi.ac.uk/hartlib/> (last accessed 7 June 2019).

¹⁰ See the brief history of attempts to assemble and edit Comenius's correspondence, with a bibliography of published editions, compiled by Vladimír Urbánek: <http://cofk.history.ox.ac.uk/reference/jan-amos-comenius/> (last accessed 7 June 2019).

¹¹ The first volume has now appeared: *Johannis Amos Comenii Opera Omnia*, vol. 26/I, *Epistulae*, Pars I, 1628-1638 (= *Dílo Jana Amose Komenského*, vol. 26/I, *Korespondence*, 1. část, 1628-1638), ed. Martin Steiner, Tomáš Havelka, Vladimír Urbánek et al. (Prague: Academia, 2018).

¹² CofK blog: 13 Oct. 2015, <http://www.culturesofknowledge.org/?p=5891> (last accessed 7 June 2019).

core data of this kind was viable and attractive for several reasons. First, all potential data providers and users readily agreed on the basic fields of a data model. Second, all of these types of data were typically recorded on autograph letters themselves and provided in scholarly inventories and editions of correspondence. Third, basic tools and standards already existed for dealing with people, places, and dates. Fourth, searching, sorting, analysing, and visualising correspondence data by these six fields offered a research programme in its own right. Fifth, core metadata could also link to richer resources elsewhere, allowing the union catalogue to network diverse resources offered by publishers, archives, and research projects.

A sixth advantage was the freedom to continue adding supplementary data fields long after the commencement of the project. For instance, some calendars record information on the physical attributes of manuscript letters, such as watermarks, seals, ribbons, or even the manner in which the letter was folded. Others include detailed information on the contents of the letter: incipits and explicits, people and places mentioned, or topics discussed. A comprehensive union catalogue should be able to accommodate all available data of this kind, without obliging contributors to provide data in any but the short list of core fields.

Work on devising such a model began in the planning phase. An appendix to the original application contained the first draft of a ‘database structure for the union catalogue’, consisting of 45 fields, compiled by Philip Beeley and Neil Jefferies. On 10 July 2009, the project’s first main technical meeting scrutinized this first draft, along with the metadata standards which Robert McNamee had developed for the Electronic Enlightenment Project.¹³ To ensure the adequacy of the model, 1,000 atypical letter records were collected for test fitting in a mock-up database constructed by Sue Burgess. Detailed feedback was provided by a specialised sub-committee.¹⁴ Richard Sharpe then played the lead role in further polishing the model and overseeing its implementation in EMLO during Phase I.¹⁵

4.2 Enhanced systems development

A second technical challenge was to design and build a digital system which allowed a team of scholars to curate tens of thousands of core and supplementary letter records. Here again, results considerably surpassed original intentions. Written by a group of scholars with little technical expertise, our original proposal lacked a technical blueprint altogether until mere days before the application deadline, when Neil Jefferies sketched out plans for catalogue storage, data import procedures, editing tools, and a website.¹⁶ Our initial budget allocated only one year of full-time development to create this basic system.

The first year revealed that we needed something much more ambitious. Created over a long period of time by many people working without clear standards, the disorderly data derived from the Bodleian card catalogue provided the perfect testbed for developing tools for data cleaning, standardisation, and reconciliation. The Steering Committee requested and the Mellon Foundation authorized a further two and one-half years of full-time systems development, funded by the exchange-rate windfall, to create far more elaborate functionality for editing, searching, and browsing than originally planned.

¹³ <http://cofk.history.ox.ac.uk/events/digital-round-tables-2009/> (last accessed 07 June 2019).

¹⁴ This sub-committee was led by Jackie Stedall (†) and Benjamin Wardhaugh and also included Kateřina Horníčková, Leigh Penman, and Helen Watt.

¹⁵ This latter task was subsequently co-ordinated by Miranda Lewis.

¹⁶ Neil subsequently acted as Technical Strategist throughout Phases I and II, as well as serving on the Steering Committee and contributing key input to the technical discussions at the heart of the COST Action.

4.3 Back End: EMLO-Edit

The most exciting and innovative systems development in Phase I took place in the ‘back end’ invisible to external users. The resulting system, ‘EMLO-Edit’, is a hand-crafted, user-friendly interface which enables a geographically distributed community of authorised data editors to add, correct, and augment information both prior to public release and (following further refinement in Phase II) after publication as well. Showing great intuition into the needs of scholars editing this messy data, our Phase I systems developer, Sushila Burgess, developed this bespoke suite of online editing tools with an SQL implementation on a LAPP (Linux-Apache-PostgreSQL, PHP) platform.

The function of these tools is best explained in terms of the problems they solve. To begin with, the Bodleian cards often referred to the same person in many different ways. Standardisation of this nomenclature was facilitated by editable and merge-able pick lists. Further tools provide a simple system for ‘merging’ people, places, and institutions which had been expressed in multiple ways,¹⁷ and for enhancing records with digital images, licensing information, and links to other resources.

A third category of tools facilitated communication within the editorial team. The innumerable editorial decisions involved in this kind of data cleaning often require pooling the specialist expertise of the entire editorial team. To expedite this, we developed a ticketing system which allowed one data editor, working within EMLO-Edit, to query a precise detail with another editor without the need for communication outside the system. This allowed each member of an editorial team to concentrate primarily on their areas of greatest expertise, simultaneously accelerating the rate of editorial work and improving its quality. Moreover, working this way within EMLO was unaffected when team members moved from the other side of the room to the other side of the world. Very early in the project history, these improvements raised our aspirations considerably: no longer content merely to create a union catalogue, a finding aid, or even a tool for analysis and visualisation, we henceforth aimed to create digital infrastructure capable of facilitating unprecedented forms of radically multilateral collaboration amongst a dispersed network of scholars, projects, repositories, and systems developers.

Collaborative editing raised obvious issues of quality control. Our first response to this challenge was to introduce a logging system into EMLO-Edit, which tracks each contributor’s every intervention by means of SQL-based audit trails. When all goes well, this arrangement allows the system to credit contributors. When things go wrong, it also permits system administrators to locate and reverse all changes made by a given contributor.

In short, thanks to the good fortune of a budgetary surplus, an engaged community of expert data editors, an insightful and responsive systems developer, and able management from the core project team, the rudimentary editorial toolset originally proposed rapidly evolved into a powerful publishing and content management system for epistolary records. This has provided the basis for our subsequent experiments in community cataloguing.

4.4 Front End: EMLO

¹⁷ The semi-automation of metadata standardisation was taken to another level by the development of the ‘Recon’ tool in Phase III, in partnership with colleagues at Aalto University in Helsinki. Further detail in section 8.2 of this chapter.

EMLO's 'front end' allows users to interrogate all the epistolary records that have been publicly released. This user interface was designed and implemented by Monica Messaggi-Kaya and Matthew Wilcoxon using Python, the Pylons framework, and Apache-SOLR indexing on a Linux foundation. As well as the default basic search, an advanced search page allows users to search across any combination of the 59 fields of the letter data model. Full-text searching of textual content (currently confined to abstracts, incipits, and explicits) returns the relevant snippet of text in the main results list. A separate search mode allows browsing by people, location, organization, repository, or year in combination with gender. Faceted listings arrange results chronologically, by sender/recipient, or by location. Crucially, the edit button on each letter record enables authorized users to move with a single click from the front end to EMLO-Edit, where full facilities for editing the letter record are available.

4.5 Landmarks and launch

A series of one-day consultations were staged throughout the development process. In February 2010, a workshop scrutinized the pilot database, tested its online editorial interface, and compared individual catalogues against project standards.¹⁸ On 22 September 2010, the prototype version of the union catalogue was showcased during the first CofK conference. In February 2011, Kim McLean-Fiander assembled a group of librarians, scholars, systems developers, web designers, and user experience specialists from outside the project, who spent the morning critiquing the catalogue's function and design and the afternoon refining the interface on a screen-by-screen basis.¹⁹ On 12 May 2011, a third focus group was assembled to help name the resource prior to launch. Melissa Terras had recently advised that that online resources 'should be taking [their] digital identity and digital presence a lot more seriously'.²⁰ 'Naming your brand,' experts added, 'is the most challenging, momentous, and necessary phase in the process of branding'.²¹ With this challenge in mind, Kim took another hand-picked group through a series of exercises designed to unlock associative thinking, which helped produce a solution a few days later.²² 'Early Modern Letters Online' not only captured the essence of the project: it also generated the acronym 'EMLO', which aligned neatly with established resources such as EEBO (Early English Books Online) and ECCO (Eighteenth-Century Collections Online), to which has more recently been added EMMO (Early Modern Manuscripts Online) at the Folger Shakespeare Library.²³ Equipped with James Brown's elegant logo and visual identity, the beta version of EMLO was launched at a

¹⁸ First use group, 2 Feb. 2010: <http://cofk.history.ox.ac.uk/training-day/index.html> (last accessed 7 June 2019).

¹⁹ Second user group, 4 Feb. 2011: <http://cofk.history.ox.ac.uk/union-catalogue-focus-group/index.html> (last accessed 7 June 2019).

²⁰ Terras, 'Present, not voting: digital humanities in the Panopticon', closing plenary, Digital Humanities 2010, King's College London; blog dated 10 July 2010: <http://melissaterras.blogspot.com/2010/07/dh2010-plenary-present-not-voting.html> (last accessed 7 June 2019).

²¹ Bill Chiaravalle and Barbara Findlay Schenck, *Branding for dummies* (Hoboken, 2007), p. 111.

²² Focus group, 12 May 2011: <http://cofk.history.ox.ac.uk/naming-the-union-catalogue/index.html> (last accessed 25 June 2019).

²³ EEBO: <https://eebo.chadwyck.com/home>. ECCO: <https://www.gale.com/primary-sources/eighteenth-century-collections-online> (last accessed 25 June 2019). EMMO: <https://emmo.folger.edu/> (last accessed 25 June 2019).

reception in the Divinity School of the Bodleian Library on 22 December 2011 and released to the public online in January 2012.²⁴

5. Community building (Phases I-II)

During Phase I, EMLO remained a resource designed, built, and populated in Oxford. Yet from the very beginning, the intention was to create the nucleus of a union catalogue that would eventually be populated collaboratively by a whole community of likeminded scholars, each benefitting from all the work – conceptual, technical, and scholarly – already contributed by the others. By creating a natural home for data in this field, we also hoped to address the problem of sustainability caused by the constant proliferation of data silos, while allowing the scholarly community to move beyond individual correspondences to study the broader networks of which they were part, and to create a more adequate showcase for work in this field.

The potential collaborators in this enterprise were many and various: doctoral students, postdoctoral researchers, mid-career scholars with large archives of unpublished data, senior scholars wishing to find a safe haven for the results of scholarly lifetimes of archival spadework, scholarly publishers, librarians, and archivists. The sources of data available to them were likewise diverse: uncatalogued manuscripts, catalogues of archival collections, early modern printed correspondence, modern inventories and critical editions, and a proliferating range of digital resources. During Phases I and II, building a community embracing all of these categories was pursued principally by means of four different kinds of activity: hosting scholarly meetings in Oxford, showcasing the project in presentations elsewhere, developing electronic media, and engaging an emerging generation of scholars.

5.1 Seminars, colloquia, workshops, conferences, and presentations

The core of these community-building efforts was a series of three annual seminars in Oxford in 2010, 2011, and 2012 which culminated in a concluding conference. The first two seminars, organised by Pietro Corsi and Peter Harrison, revealed an unexpectedly keen appetite for discussion of learned correspondence within the local community. Instead of developing into a fairly specialised affair patronised mainly by project members, the seminar consistently overflowed a standard seminar room, necessitating a move to the History Faculty's lecture theatre. Equally important were the informal discussions which continued during the wine receptions following each paper. A conference organised by Noel Malcolm and Rhodri Lewis in September 2012 brought many internationally leading experts to Oxford to discuss the epistolary exchange of learning in early modern Europe all its diverse forms.²⁵

Parallel to this core series, a second sequence of events reached out from Oxford to east-central Europe. Thematically, this series was more narrowly focussed on the ethos and networks of Protestant refugees from the Thirty Years War, which featured so prominently in the circles of Hartlib and Comenius. Geographically, it was more wide-ranging. The first colloquium met in Prague in January 2009 to discuss the eschatological expectation of these

²⁴ Beta launch, 22 Dec. 2011: <http://cofk.history.ox.ac.uk/early-modern-letters-online-beta-launch-event/index.html> (last accessed 25 June 2019).

²⁵ 'Communities of Knowledge: Epistolary Cultures in the Early Modern World', Faculty of English, University of Oxford, 20-22 Sept. 2012: <http://communitiesofknowledge.history.ox.ac.uk/> (last accessed 25 June 2019).

refugee groups.²⁶ The second met in Cracow in July to discuss educational reform and ecclesiastical reconciliation.²⁷ The third convened in Budapest in April 2010 and was devoted to encyclopaedism, pansophia, and universal communication.²⁸ An omnibus conference in Oxford in September 2010 on the topic of ‘Universal Reformation: Intellectual Networks in Central and Western Europe, 1560-1670’ brought many of the results of this series together.²⁹ A further conference in September 2011, attended by several of the other projects discussed in this volume, aimed to articulate a conceptual and methodological framework suitable for data-driven analyses of the complex ‘intellectual geography’ especially of war-torn central Europe in this period.³⁰

A third strand of meetings inaugurated in Phase I was technical in nature. CofK’s first international meeting was a technical workshop held in St Anne’s College, Oxford, in March 2010. One purpose of the meeting was introductory: ten rapid presentations of major correspondence projects and collections in the Czech Republic, Denmark, Germany, Italy, the Netherlands, the UK, and the US.³¹ Its deeper purpose was to initiate an international discussion of the issues involved in creating metadata standards and interoperable technical systems for the collection, storage, and exchange of epistolary data and metadata, with a view to enhancing flexibility, scalability, sustainability, and opportunities for collaborative work. In retrospect, this meeting outlined much of the agenda of the subsequent COST network between 2014 and 2018 (discussed below in section 8.1) and assembled for the first time many of the participants in it.

In the meantime, the fourth and fifth seminars, in 2013 and 2014, continued the discussions by bringing to Oxford representatives of some of the leading projects employing innovative digital tools and methods to explore the epistolary cultures of the early modern and

²⁶ ‘Apocalypticism, Millenarianism, and Prophecy: Eschatological Expectations between East-Central and Western Europe, 1560–1670’, 15–16 January 2009, Institute of Philosophy, Academy of Sciences of the Czech Republic, Prague; organised by Vladimír Urbánek and Howard Hotson: <http://cofk.history.ox.ac.uk/events/prague-2009/> (last accessed 25 June 2019).

²⁷ ‘Educational Reform, Philosophy, and Irenicism, 1560–1670’, 6–8 July 2009, Polish Academy of Arts and Sciences, Cracow; organised by Vladimír Urbánek, Howard Hotson, and Steffen Huber: <http://cofk.history.ox.ac.uk/events/krakow-2009/> (last accessed 25 June 2019).

²⁸ ‘Encyclopaedism, Pansophia, and Universal Communication, 1560–1670’, 14–16 April 2010, Semmelweis Museum, Library, and Archives of the History of Medicine; and Central European University, Budapest; organised by Márton Szentpéteri, Gábor Kecskeméti, Benedek Varga, and Márton Zászkaliczky: <http://cofk.history.ox.ac.uk/events/budapest-2010/> (last accessed 25 June 2019).

²⁹ Organised by Howard Hotson and Vladimír Urbánek, St Anne’s College, 21–23 Sept. 2010: <http://universalreformation.history.ox.ac.uk/> (last accessed 25 June 2019).

³⁰ ‘Intellectual Geography: Comparative Studies, 1550–1700’, organised by Howard Hotson, St Anne’s College, Oxford, 5–7 Sept. 2011: <http://intellectualgeography.history.ox.ac.uk/> (last accessed 25 June 2019). The speakers at this conference included representatives of MtRoL and FBTEE, see chapters 3 and 7 of this volume).

³¹ In addition to CofK, these included ‘Circulation of Knowledge and Learned Practices in the Seventeenth-Century Dutch Republic: A Web-Based Humanities Collaboratory on Correspondences’ (Descartes Centre/KNAW/Huygens Institute); ‘Mapping the Republic of Letters: From Individual Scholarship to Large-Scale Collaboration’ (Stanford Humanities Center); ‘Topographie der Gelehrtenrepublik’ (Forschungsbibliothek, Gotha); the Centre for Editing Lives and Letters (QMUL); a preview of the Galileo//thek (Museo Galileo, Florence); and four major libraries with large collections of early modern learned letters: The Royal Library, Copenhagen; the British Library, London; National Museum Library, Prague; and the Herzog August Bibliothek, Wolfenbüttel. Further detail at <http://cofk.history.ox.ac.uk/events/digital-round-tables-2010/> (last accessed 25 June 2019).

other periods.³² These events were complemented by over 40 presentations of the project in the UK³³ and nine EU countries as well as Israel, Canada, and the United States.³⁴

5.2 Digital media

Community creation was assisted by the innovative work of CofK's first Project Manager, James Brown, in devising digital media for the project. Thanks to remarkable graphic design talents, James's core responsibility for event planning spilled over into the production of beautiful posters and flyers, slide presentations, and logos for both CofK and EMLO which are still used almost a decade later. James also rapidly emerged as an accomplished web designer with a sharp sense of emerging best practice and breaking trends: he was amongst the first project coordinators in our field, for instance, to assemble comprehensive galleries, first of podcasts, and then of videos of seminars and conferences.³⁵ James also built microsites for each of the three conferences hosted in Phase I, complete with speakers' biographies, paper abstracts, podcasts, videos, and blogs. Communications were disseminated from the outset via an email list with over 1,000 subscribers and an RSS feed. In the teeth of scepticism from the Steering Committee, James opened a Twitter account in May 2011 which has now attracted over 2,500 followers.³⁶

The most extensible aspect of the website, and perhaps its most effective conduit of communication, is the project blog. Initiated on 19 April 2009 – only three months into the first grant period – the CofK blog has provided a running commentary on the project's

³² Abstracts and podcasts of the 11 papers are available at http://www.culturesofknowledge.org/?page_id=1270 and http://www.culturesofknowledge.org/?page_id=4861 (last accessed 25 June 2019). These seminars also provided first contact with a number of future collaborators, including Huib Zuidervaart of the Huygens; Christoph Kudella (then of University College Cork, now of Göttingen); Eero Hyvönen of Aalto University (see section 8.2); and Ruth and Sebastian Ahnert (see section 8.5)

³³ KCL, 2010; CERL Advisory Task Group, March 2011; the universities of Plymouth, April 2011, and Reading, July 2011; British Library, April 2011 and July 2012; UCL, Sept. 2011; Cambridge, June 2012; The Royal Society, September 2012; AHRC Digital Transformations Moot, London, Nov. 2012; Durham, July 2013.

³⁴ Canada: University of Saskatchewan, April 2012. Czech Republic: Institute of Philosophy, Academy of Sciences, Prague, Nov. 2012, Sept. 2013. France: Institut protestant de théologie, Paris, April 2013; Observatoire, October 2013; Sorbonne, Dec. 2013. Germany: Berlin-Brandenburg Academy of Sciences, September 2010; Forschungszentrum Gotha, April 2012; HAB, Wolfenbüttel, Sept. 2012; Heidelberger Akademie der Wissenschaften, Sept. 2013. Hungary: CEU, Budapest, April 2010. Israel: National Library, October 2013. Italy: International Society for Intellectual History Conference, Verona, May 2009; CINI Foundation, Venice, March 2011; EUI, Florence, Nov. 2013. The Netherlands: Huygens ING, The Hague, June 2011; Utrecht, June 2013. Poland: University of Warsaw, October 2010. Sweden: University Library, Uppsala, June 2011; National Library, Stockholm, June 2011; Gothenburg, Sept. 2011. Switzerland: Université de Lausanne, Nov. 2013. United States: Redwood City, California, May 2011; Stanford University, August 2012.

³⁵ Three dozen podcasts of the seminar papers between 2010 and 2014 (many complemented by slides and links) are available here: http://www.culturesofknowledge.org/?page_id=187 (last accessed 25 June 2019). Two dozen videos of papers from the 2011 conference are available here: http://www.culturesofknowledge.org/?page_id=189 (last accessed 25 June 2019).

³⁶ <https://twitter.com/cofktweets>.

development currently consisting of over 200 posts.³⁷ During the project's infancy in 2009, its main function was to help consolidate a community by disseminating news of cognate projects.³⁸ CofK's own project news – opportunities, appointments, presentations, and members' activities – was gradually stirred into the mix. From the spring of 2010 onward, the blog also featured detailed accounts of seminar papers, workshops, and conferences. Following the relaunch of EMLO in January 2015 (see section 6.3), the blog came to be dominated by Miranda's gradually accelerating announcements of new catalogues, interspersed with news and opportunities arising from the COST Action (section 8.1).

5.3 Engaging early career scholars

During Phase I, these community-building initiatives revealed a keen appetite to contribute data to a common resource but also disclosed a practical obstacle: many established scholars were prevented from undertaking the meticulous editing necessary to prepare their material for publication in EMLO by the multiple competing obligations arising from their academic posts. What they needed was access to emerging scholars able to edit data for publication under supervision. To meet this need, Phase II included funding for a network of trained 'Digital Fellows'. To selected early-career scholars, the DF programme offered training in EMLO's tools and standards, the opportunity to prepare material for publication within a major collaborative research project, and hourly-paid work at a pace adapted to their needs. To more senior scholars the programme offered access to the kind of reliable and well-trained assistants needed to polish their material for publication. To CofK the programme offered a flexible and cost-effective means of expanding the community of people working actively within EMLO while accelerating the rate at which new catalogues could be prepared for publication.

Miranda Lewis's role grew from EMLO Editor to network-builder in her own right, with a lead role in recruiting data contributors and matching them up with Digital Fellows. During Phase II, the number of DFs on her books grew from zero to 13, and these were soon joined by a number of other categories of early career collaborators funded from their home countries or institutions.³⁹

³⁷ The blog on the first website is at <https://cofk.history.ox.ac.uk/office-of-addresse/index.html> (last accessed 25 June 2019). The second is at http://www.culturesofknowledge.org/?page_id=6 (last accessed 25 June 2019).

³⁸ Those institutions featuring in more than one blog during Phase I included The Royal Society (1 Sept. and 20 Nov. 2009, 14 July and 29 Oct. 2010, 1 Jan. and 14 May 2012); The Centre for Editing Lives and Letters; QMUL / UCL (1 Sept. 2009, 12 Oct. 2012, 14 Dec. 2012); the Electronic Enlightenment (10 Dec. 2009, 26 Oct. 2010, 2 Aug. 2011); the Scaliger Institute, Leiden (1 Dec. 2010; 21 Jan, 15 March, and 4 Oct 2012); the Reading Conference in Early Modern Studies (12 Dec 2010, 22 July 2011); the Edward Worth Library (14 Feb 2011, 22 Feb 2012); and the Circulation of Knowledge project at the Huygens ING (7 June 2011, 15 March 2012, 29 August 2012). Other links were celebrated with Warsaw (21 May 2010), Stanford (21 March 2011), Scientiae (18 May 2011), Utrecht (19 Sept 2011), Prague (11 Jan 2012), Warwick (26 March 2012); the Maison Française d'Oxford (17 July 2012), *The Seventeenth Century* (28 Sept 2012), the Warburg Institute (18 Dec 2012), and Durham (30 Nov 2012).

³⁹ These include Marc Kolakowski, who spent the academic year 2014–2015 in Oxford, funded by a Doctoral Mobility Study Grant from the Swiss National Science Foundation, working on a network of correspondence relating to the Zürich theologian and churchman, Johann Wilhelm Stucki (1542–1607); Charlotte Marique, who while studying at the Université de Liège, spent four months as an intern in the spring of 2014, before becoming Miranda's trusted Editorial Assistant from 2017; and Cristiano

Since 2016, CofK has also helped launch a number of educational pilot projects at the undergraduate as well as graduate level. One such initiative, known as ‘Bodleian Student Editions’, is a series of scholarly editing masterclasses arising from a collaboration between CofK and the Bodleian. Working from both original manuscripts and digital facsimiles, participants catalogue and transcribe unpublished letters from Bodleian collections selected by curators. In the process they are introduced to a range of skills including the proper handling of special collections, palaeography, metadata creation and analysis, textual mark-up and encoding, and scholarly editing for print and digital media.⁴⁰ Similar initiatives are now being introduced into mainstream credit-bearing papers in the History Faculty, such as Giora Sternberg’s second-year, source-based paper, ‘Writing in the Early Modern Period, 1550–1750’.⁴¹ These activities have introduced over 100 active contributors to EMLO, with many more currently preparing material for publication.

6. Re-engineering EMLO Mach II (Phase II: Jan 2013-March 2015)

The community-building activities of Phase I provided an obvious and ambitious systems-development agenda for Phase II. In order to transform EMLO into a platform for multilateral collaboration, major technical upgrades were needed to both back- and front-end systems. In the back end, we needed expedited procedures for ingesting large datasets in bulk, a webform for users to input their data directly into the system, and multiple independent user accounts, to allow scholars in and outside the project team to curate their data. In the front end, we needed new means of acknowledging contributors and describing their contributions. The intrinsic difficulty of this work was exacerbated by staffing changes. In early 2013, the Lead Developer in Phase I, Sushila Burgess, left the project. Her replacement, Renhart Gittens, was unable to begin full-time work immediately and left at short notice in September 2014. Mat Wilcoxson then stepped into the Lead Developer role from 2014 to 2019.

6.1 Re-engineering the back end: EMLO-Edit

The first task was to semi-automate workflows for ingesting catalogue data in bulk. The original version of EMLO had been constructed to curate a small number of datasets created in house. During Phase I, new data sets arrived in a great variety of file formats, some now obsolete. Ingesting such diverse datasets required laborious, ad-hoc data-manipulation by the systems developer or data analyst. To address this problem, Renhart worked with Miranda to create an ingestion template, standardize an abbreviated upload procedure, and automate several processes previously performed manually. Thanks to this work, an unproblematic spreadsheet can now be uploaded in a few minutes without technical assistance, and a simple user

Amendola and Antonio Geremicca have help foster links with the EpistolART project directed by Paola Moreno in Liège, <http://www.epistolart.ulg.ac.be/> (last accessed 25 June 2019); Cristina Erquiaga (a doctoral student in Salamanca) and Francesco Zambonin (on an undergraduate exchange from the Collegio Galileo in Padua), who helped increase the profile of Spanish and Italian scholars on EMLO.

⁴⁰ The origins of the project is recounted at <https://www.e-pistolary.net/speaking-in-absence/explore-bodleian-collections/> (last accessed 25 June 2019). The CofK blog for 12 Dec. 2016 announced the first of these editions. The catalogue page for the series is <http://emlo-portal.bodleian.ox.ac.uk/collections/?catalogue=bodleian-student-editions> (last accessed 25 June 2019).

⁴¹ Cf. the CofK blogs for 1 March 2016, 30 July 2017, and 2 March 2018.

interface allows the Digital Editor to process more complicated data sets without developer intervention.

A second main task in Phase II was to provide tools to help users input their data in a form pre-validated for publication in EMLO. The initial plan was to provide separate tools for preparing data off-line as well as online. Yet by 2014 this project was rendered obsolete by two simultaneous developments. On the one hand, as EMLO became more sophisticated, the speed and reliability of the tool suffered, since it needed to process increasing quantities of data to keep pace. On the other, as wifi access became increasingly ubiquitous, the need for an offline tool rapidly diminished.

Our priority now became the development of an online EMLO interface to hold data in a private, secure working area linked to the main EMLO-Edit database. This interface has three main advantages. First, it is capable of validating and reconciling incoming data with reference to all the existing people and place records on EMLO. This ratification, second, increases the efficiency of data entry, greatly reducing the time required to input a letter record. Third, it also eliminates the need for data to be manipulated in a secondary upload stage by the EMLO Editor.

This second task was closely related to a third and even more complex and crucial one: to create individual user accounts in which editors in and outside the project team could curate data prior to publication. The original EMLO backend was designed to support a small project team of paid researchers and editors, all of whom had access to and editorial rights for all the 60,000 records on the system, which were only published when the entire enormous dataset was ready. Major alterations were needed to provide adequate infrastructure for collaborative work on a far broader scale. Contributors needed to be able to curate their data in a private space, in the context of all the person, place, and repository records shared across catalogues. Yet it was equally important that contributors should *not* be able to edit other data on the system. The most complex technical problem EMLO has yet faced was to re-engineer the data entry system to allow rapid integration of new data with all the records in EMLO-Edit while keeping that new data private and outside EMLO-Edit until all editorial checks had been completed and contributions were ready for publication.

6.2 Re-working the front end: user interface and catalogue pages

These back-end changes to EMLO were complemented by an overhaul of the front end. In autumn 2014, Monica Messaggi-Kaya redesigned the aesthetics of EMLO's front end, while Mat Wilcoxson added new functionality. Separate profile pages for works, people, and places were created, featuring enriched metadata, images, links to related resources, and simple visualizations. EMLO was also made mobile- and tablet-ready.

The crucial innovation, however, was to provide each catalogue in EMLO with its own page providing information on the correspondence in question, the principle correspondent, the source and provenance of the data, comprehensive scholarly credits, and acknowledgement of funding and partnerships. Meanwhile, every individual letter record now both identified the source of the record and linked back to the catalogue page for more detail. In short, this new system, with its careful delineation of contributor controls, editorial responsibilities, private working spaces, and prominent acknowledgement of contributions, transformed EMLO from the nucleus of a union catalogue to a platform for multilateral scholarly collaboration.

6.3 New data and the re-launch of EMLO, December 2014-January 2015

Alongside these major technical overhauls, the final main task of Phase II was to prepare for publication a first series of catalogues contributed by extra-mural partners. The choice of dataset to showcase in the relaunch was obvious. From the first year of the project onward, a series of bi-lateral exchanges had developed a close working relationship with the ‘Circulation of Knowledge’ (CKCC) project in the Huygens Institute for the History of the Netherlands.⁴² This large project had digitized over 20,000 letters to and from nine figures central to the Dutch golden age and had published them in the ePistolarium in 2013 as the basis for a pioneering experiment in natural language processing and topic modelling of multilingual text corpora. Nothing could have been better suited to inaugurating a new phase of work than the publication of this huge new dataset.

Thanks to these developments, Phase II culminated in a relaunch of EMLO in December 2014-January 2015.⁴³ This was a milestone event for several reasons. First, publication of 20,000 new records increased the size of EMLO by one third.⁴⁴ Second, these new records represented a single, large, and coherent new cluster: nine new catalogues relating to the Dutch golden age.⁴⁵ Third, the contribution of this huge dataset by a major new partner represented a breakthrough in CofK’s aspiration to create a catalogue populated through large-scale international collaboration. Fourth, each new letter record linked back to the full letter texts on the ePistolarium, confirming EMLO’s potential to become a gateway to richer resources held by partner projects. Fifth, the redesigned front-end permitted the prominent attribution of each dataset to its contributor; and the re-engineered back end allowed the simultaneous curation of multiple catalogues prior to publication. Facilitated by these developments, by March 2015 44,000 new letter records were in active preparation in the newly created private user accounts. Sixth and perhaps most important of all, these innovations permitted the onset of a rolling programme of publication of new datasets.

Guiding the project through to this major landmark was the culminating achievement of Elizabeth Williamson. After helping to coordinate the first main event in the COST Action (see section 8.1), Lizzy stepped down in April 2015,⁴⁶ and was succeeded as Project Manager for Phase III and IV by Arno Bosse.⁴⁷

7. Populating a union catalogue collaboratively: Nov 2014 – Sept 2018

⁴² Martin Bossenbroek introduced CKCC in Oxford in July 2009. Howard Hotson, Neil Jefferies, and James Brown provided a status update on EMLO at the Huygens ING in June 2011. Charles van den Heuvel participated in the Intellectual Geography conference in September 2011 and Howard Hotson helped launch of the ePistolarium at Utrecht University in June 2013. On the project, see <http://ckcc.huygens.knaw.nl/> (last accessed 25 June 2019).

⁴³ CofK blog: 16 Jan. 2015.

⁴⁴ By the end of Phase II (31 March 2015), 87,417 letter records were available to the public, along with 19,936 people records, and 5,044 place records in 24 catalogues.

⁴⁵ Caspar Barlaeus (1584–1648), Isaac Beeckman (1588–1637), René Descartes (1596–1650), Hugo Grotius (1583–1645), Christiaan Huygens (1629–1695), Constantijn Huygens (1596–1687), Anthonie van Leeuwenhoek (1632–1723), Dirck Rembrantsz van Nierop (1610–1682), and Jan Swammerdam (1637–1680).

⁴⁶ CofK blog: 24 July 2015.

⁴⁷ CofK blog: 20 April 2015.

7.1 Community-building and collaborative population

This re-engineering of EMLO finally created a suitable environment for CofK's central experiment in populating a union catalogue of the republic of letters collaboratively. This intention had been publicly announced at the outset of the project in Verona in May 2009.⁴⁸ At that point, the nascent EMLO project was a highly speculative digital offshoot of a basically hard-copy editorial project. The speaker recalls a distinct sense of unease in issuing an invitation to a new kind of gathering which no one might wish to attend. Such apprehension, it turns out, was unjustified.

Only two weeks into the project – at the inaugural colloquium in Prague on 15-16 January 2009 – Noémi Viskolcz (Szeged) offered EMLO its first donation: her published catalogue of the correspondence of Johann Heinrich Bisterfeld (1605-1655), and transcriptions of the correspondence of Johann Permeier (1597–1644?).⁴⁹ On 9 March 2010, at the first international meeting convened by CofK, Gillian Bepler proposed to press ahead with longstanding aspirations at the Herzog August Bibliothek in Wolfenbüttel to catalogue the huge correspondence of Johann Valentin Andreae' (1586–1654); several years later the DFG awarded a grant for an inventory and digital archive of Andreae's nearly 5000 letters.⁵⁰ The following month, a conference on John Selden (1584-1654) co-sponsored by CofK, resulted in another valuable contribution: G.J. Toomer's meticulous edition of the letters of John Selden.⁵¹

The first CofK seminar, organised by Pietro Corsi and Peter Harrison in May and June of 2010, resulted in further contributions.⁵² Of six speakers who spoke on individual correspondents, five have now contributed large correspondence catalogues to EMLO: namely, those of Joseph Scaliger (1540-1609),⁵³ Antoinette Bourignon (1616–1680),⁵⁴

⁴⁸ Annual conference of the International Society for Intellectual History, 26 May 2009

⁴⁹ CofK blog: 26 March 2016.

⁵⁰ First instalment announced in CofK blog: 6 July 2015; cf. <http://diglib.hab.de/?link=042> (last accessed 25 June 2019).

⁵¹ 'John Selden, 1584-1654: Scholarship in Context', Magdalen College, Oxford, 24-26 June 2010, organised by Thomas Roebuck (now the University of East Anglia) and Jeffrey Miller (now at Montclair State University): <http://www.cems.ox.ac.uk/selden/> (last accessed 25 June 2019).

⁵² <https://cofk.history.ox.ac.uk/events/seminars-2009-10/> (last accessed 25 June 2019).

⁵³ Contributed as a working spreadsheet by Dirk van Miert (Huygens Institute, now Utrecht) and Paul Botley (Warburg Institute, now Warwick), this was the first catalogue to appear after the relaunch of EMLO. The publication of the underlying edition was celebrated by a reception in the Divinity School on 12 September 2012 featuring celebratory remarks by Anthony Grafton (Princeton), whose Balzan Prize for 2002 had co-funded (with the Mellon Foundation) the creation of the edition at the Warburg Institute in London and the Scaliger Institute in Leiden: <http://cofk.history.ox.ac.uk/correspondence-of-joseph-justus-scaliger-published/index.html> (last accessed 25 June 2019).

⁵⁴ Contributed by Mirjam de Baar (Groningen, now Leiden), see CofK blog, 6 Dec. 2015.

Nicolas-Claude Fabri de Peiresc (1580–1637),⁵⁵ James Ussher (1600–1656),⁵⁶ and Francis Bacon (1561–1626) – the last of which is still in preparation.⁵⁷

The time-lag between seminar and publication, however, has been considerable. All of these contributions had to await the overhaul of EMLO's infrastructure in 2015, and some needed to wait still longer until the moment was right for the contributors. The length of these intervals demonstrates a fundamentally important point: the creation of a union catalogue populated by the scholarly community is inevitably a long-term undertaking. In addition to creating tools and infrastructure, it involves providing proof of concept, cultivating long-term relationships with potential contributors, and adapting every aspect of the publication platform to meet their needs.

7.2 Dimensions of expansion

In choosing which catalogues to pursue and in what order, the project team and Steering Committee were conscious of a large number of competing criteria. The quickest way to enlarge EMLO was to upload large bodies of readily available data, but we also prioritised previously unpublished material. Although basic letter records were quicker to process, we sought to attract catalogues with enriched metadata, accompanying images and texts, or links to other resources elsewhere. A wide range of contributors was also desirable: doctoral students, retired colleagues, partner projects, publishers, and archives. Thematically, we needed both to build out from our initial centres of interest and to publish data on earlier and later periods, other areas, and different communities to add scope of the resource.

One obvious strategy was gradually to expand the project's two original clusters of correspondence. The original calendars of the letters of early Fellows of the Royal Society (Aubrey, Lhwyd, Lister, and Wallis), supplemented by data from the ePistolarium (Huygens and Leeuwenhoek), were incrementally enhanced by the addition of Elias Ashmole (1617–1692), Henry Oldenburg (c.1619–1677), Robert Boyle (1627–1691), John Collins (1625–1683), Francis Vernon (1637–c. 1677), and Robert Plot (1640–1696), as well as later figures such as James Jurin (1684–1750), Thomas Pennant (1726–1798), and Charles Hutton (1737–1823). A recent pilot grant from Oxford's Fell Fund has allowed the addition of the 4,390 'early letters' passed to EMLO from the Royal Society's own archive, and the correspondences of John Locke (1632–1704), Isaac Newton (1642–1726/7), John Flamsteed (1646–1719), and Edmond Halley (1656–1742).⁵⁸

⁵⁵ This correspondence posed significant challenges to assemble. Peter Miller (Bard College) contributed metadata from the Tamizey de Larroque edition, and CofK added records from de Waard's edition of Mersenne's letters to create a catalogue (published in September 2016) of 1,843 letters linked to the letter texts on Gallica; but thousands of additional manuscript letters in Carpentras and Aix still await similar attention.

⁵⁶ The three-volume edition of *The Correspondence of James Ussher* (1600–1656) by Elizabeth Anne Boran (Edward Worth Library, Dublin) appeared in 2015. The catalogue was launched during the final COST conference in Malta in February 2018.

⁵⁷ The calendar of Bacon's correspondence by Alan Stewart (Columbia and QMUL) is being prepared for publication at the time of writing.

⁵⁸ As well as the Royal Society itself, these catalogues have inaugurated partnerships with the Ashmolean Museum (Plot was its first Keeper, and the Ashmole catalogue was published to celebrate the 400th birthday of its founder in April 2017), the Electronic Enlightenment (Boyle), the Newton Project (Newton), the Curious Travellers project (Pennant), as well as Andrea Rusnock (Jurin), Alexandre J. Tessier (Vernon), and Benjamin Wardhaugh (Hutton).

To the original pair of Hartlib and Comenius were likewise added catalogues of many of their correspondents: in addition to Andreae, Bisterfeld, Bourignon, Boyle, Descartes, Oldenburg, Permeier, and Ussher, these include Joseph Mede (1586–1638), Joachim Jungius (1587–1657), Abraham von Franckenberg (1593–1652), Johannes Coccejus (1603–1669), John Winthrop (1606–1676), and John Worthington (1618–1671). The new AHRC-funded project, ‘Networking Archives’ (section 8.5 below) is now adding to this list.

To the ePistolarium’s core correspondence relating to the Dutch golden age, we have added Scaliger, Jan Gruter (1560–1627), Johannes Isacius Pontanus (1571–1639), Gerardus Joannes Vossius (1577–1649), Henricus Reneri (1593–1639), Anna Maria van Schurman (1607–1678), Isaac Vossius (1618–1689), Pierre Bayle (1647–1706), Hadriaan Beverland (1650–1716), Jean Le Clerc (1657–1736), Guilielmus Surenhusius (c. 1664–1729), and Adriaan Reland (1676–1718).⁵⁹ The catalogues of van Schurman and Surenhusius inaugurate a partnership with the Dirk van Miert’s important ERC-funded project on ‘Sharing Knowledge in Literary and Learned NETWORKS’ (SKILLNET).⁶⁰

A complementary set of learned correspondences from the Spanish Netherlands have also begun to appear. These include catalogues of 541 letters of the great cartographer Abraham Ortelius (1527–1598), contributed by Joost Depuydt (FelixArchief, Antwerp); 940 letters of the great painter-diplomat, Peter Paul Rubens (1577–1640), extracted from existing editions;⁶¹ 1,515 letters of the leading printer, Christophe Plantin (c. 1520–1589);⁶² and a prospective catalogue of his friend, the great Flemish humanist Justus Lipsius (1547–1606), currently being emended by Jeannine De Landtsheer (Leuven). For sixteenth-century Spain, Juan Luis Vives (1493–1540) and Hernán Núñez de Toledo y Guzmán (1475–1553) have recently been joined by Antonio Agustín (1517–1586), inaugurating a collaboration with the Spanish Republic of Letters project led by Guy Lazure (Windsor, Ontario);⁶³ while Benito Arias Montano (1527–1598), contributed by Antonio Dávila Pérez (Cádiz), is an early arrival stimulated by the COST Action.

With exceptions such as these, this strategy of gradual expansion has created a core data set still centred primarily on the North Sea in the seventeenth century. To expand the geographical and chronological scope of EMLO, a parallel strategy has been to develop thematic clusters. A first step was to complement the original Hartlib calendar with those of four other key seventeenth-century intelligencers. Athanasius Kircher (1602–1680) – published in February 2015, the third new catalogue to appear after the relaunch – inaugurated collaboration with another key partner: the Mapping the Republic of Letters Project in Stanford.⁶⁴ Marin Mersenne (1588–1648) was added shortly thereafter: the core metadata from the Parisian edition of his letters appeared in March, and eleven further letters published by Sir Noel Malcolm were added in September. Henry Oldenburg will be handled similarly:

⁵⁹ Although most of these catalogues were compiled with CofK funding and personnel, thanks are also due to Robin Buning (Reneri and Isaac Vossius), Karen Hollewand (Beverland), Mario Sina and Maria Grazia Zaccane-Sina (le Clerc), and Antony McKenna (on behalf of the editors of the *Correspondance de Pierre Bayle*).

⁶⁰ Cf. CofK blogs: 3 May 2018 and 25 June 2019, and SKILLNET at <https://skillnet.nl/> (last accessed 25 June 2019). Tobias Winnerling’s catalogue of Reland was also pre-processed by SKILLNET for EMLO.

⁶¹ CofK blog, 22 Apr. 2016.

⁶² CofK blog, 27 June 2016.

⁶³ Cf. CofK blog: 23 Nov. 2018, and <http://cdigs.uwindsor.ca/srl/> (last accessed 25 June 2019)

⁶⁴ Cf. CofK blogs for 10 Feb. 2015, 30 Aug. and 30 Sept. 2018, and <http://republicofletters.stanford.edu/> (last accessed 25 June 2019).

metadata from the Hall's edition was published incrementally from December 2015 onward and will be followed by records of perhaps 100 additional letters which have subsequently come to light. Nicolas-Claude Fabri de Peiresc (1580–1637) is also being handled incrementally: the metadata from the printed edition by Tamizey de Larroque edition, collated by Peter N. Miller, has been added to the extensive exchanges with Mersenne, but the vast quantity of unpublished manuscript letters remain to be added.

A related strategy was the series of disciplinary clusters launched in October 2015 with the publication of a calendar of Pierre Fermat (1607/8–1665). Added to Wallis, Huygens, and Descartes, it formed a mathematical cluster which now also includes Collins, Hutton, Giovanni Antonio Magini (1555–1617), Matthias Bernegger (1582–1640), Leonhard Euler (1701–1783), and Pietro Mengoli (1626–1686).⁶⁵ Similar clusters are being prepared for other disciplines, including astronomy, cartography, diplomacy, medicine, philology, philosophy, and theology.

Amongst the categories of contributors, early career scholars merit particular attention. Oxford doctoral dissertations have produced catalogues of Beverland (by Karen Hollewand) and of the Dublin Medico-Philosophical Society (1756–1784, by Rachael Scally). Two former CofK post-docs have made further contributions of their own: Iva Lelková on Amandus Polanus of Polansdorf (1561–1610) and Philipp Jakob Sachs von Löwenheim (1627–1672), and Robin Buning on Henricus Renneri (1593–1639) and Isaac Vossius (1618–1689). Post-doctoral scholars in and outside Oxford have done likewise, including Victoria van Hynning (letters of the recusant Tixall family, 1617–1703), Francesco Barreca (of the Museo Galileo in Florence, who refined the catalogue of Johannes Kepler, 1571–1630), and Alexandre J. Tessier (on the English scholar-diplomats, John Dodington, 1628–1673, and Francis Vernon, 1637–c. 1677).

Publishers also quickly became contributors. Librairie Droz in Geneva pioneered this form of collaboration in relation to the Scaliger calendar. Subsequent contributors include Brill (Lister), Éditions du CNRS (Mersenne), Olschki (Le Clerc, Mengoli), the Oxford University Press (Richard Baxter, Wallis), its digital wing, Oxford Scholarly Editions Online (Ashmole, Hobbes, Philip Sidney, Elizabeth Stuart, Mary Wortley Montagu), and The Voltaire Foundation (Bayle, Françoise de Graffigny, 1695–1758). The potential from this source alone is staggering: the Voltaire Foundation has published tens of thousands more letters, including the 20,000 to and from Voltaire himself, and the OSEO contains another 20,000 from the sixteenth to eighteenth centuries, not to mention the 60,000 letters in *Electronic Enlightenment* (see Chapter 2 of this volume).

Archives of correspondence not centred on individuals represent another source of data. A pioneering instance is a collection of 290 letters documenting the studies of Bohemian and Moravian students at Protestant universities in Western Europe between 1573 and 1634.⁶⁶ A second example is J. H. Hessels's inventory of several thousand letters preserved in the archives of the Dutch Church at Austin Friars in London, an important hub of Anglo-Dutch intellectual, religious, and economic commerce.⁶⁷ A rather different example is the 4,000

⁶⁵ CofK blog: 29 Oct. 2015.

⁶⁶ František Hrubý and Libuše Urbánková-Hrubá, eds., *Étudiants tchèques aux écoles protestantes de l'Europe occidentale à la fin du 16e et au début du 17e siècle* (Brno, 1970). CofK blog: Nov 2015. This collection was prepared for EMLO in partnership with the Digital Library of the Faculty of Arts, Masaryk University, Brno, where Hrubý was professor.

⁶⁷ J. H. Hessels, ed., *Epistolae et Tractatus cum Reformationis tum Ecclesiae Londino-Batavae Historiam Illustrantes: Ecclesiae Londino-Batavae Archivum*, tomus Secundus: 1544–1622 (Cambridge, 1889–97). CofK blog, 30 April 2016.

manuscript letters to and from eminent early modern Italian scholars assembled by Bartolomeo Gamba (1766–1841), subsequently divided between three different institutions: the Österreichische Nationalbibliothek in Vienna, the Biblioteca Museo Civico in Bassano del Grappa, and the Biblioteca Nazionale Marciana in Venice, where Gamba was sub-keeper. In order to reassemble it virtually on EMLO, Vittoria Feola (Padua) has assembled a team of collaborators at all of these institutions, and funding from the University of Padua (DiSSGeA) and the Medical University of Vienna, with supplements from CofK and COST Action IS1310.⁶⁸

Even further removed from the mainstream is the ‘Signed, Sealed and Undelivered’ project, which studies a collection of some 2,600 letters, sent to The Hague between 1689 and 1706, and preserved in a postmaster’s trunk now in the Museum voor Communicatie. Because the intended recipients were deceased, absent, or unwilling to pay postal fees on receipt, these letters remained undelivered, uncensored, unedited, and – in hundreds of cases – unopened. This unique collection will allow an international team to reconstruct many previously invisible features of the European postal service c. 1700, such as the folding patterns and closing techniques that built varying levels of security into these letters. New metadata standards have been developed to record in EMLO the series of physical manipulations needed to turn a sheet of paper into a folded letter enclosed within itself.⁶⁹

7.3 Specialist user communities: the case of WEMLO

No less innovative was a project conceived in Phase I: Women’s Early Modern Letters Online or WEMLO. Letters – the most ubiquitous form of writing by women surviving from the early modern period – are a vital means of recapturing women’s voices, perspectives, and agency. EMLO’s first digital editor, Kim Mclean-Fiander, was a specialist on early modern English women’s writing; James Daybell, a speaker at CofK’s second seminar, had accumulated metadata on thousands of sixteenth and seventeenth-century women’s letters;⁷⁰ another 2000 were already available in the Bodleian records which Kim was editing; and CofK’s expanding community was potentially a source for many more.

In 2012, a series of meetings, funded by the British Academy, revealed the need to combine the best of two different worlds: the study of women’s letters required a room of its own, yet that space needed to be open for easy communication with adjoining domains, to avoid creating a scholarly ghetto. What was needed, in other words, was a union catalogue within a union catalogue. The larger room – where women’s letters could be searched, browsed, sorted, and interlinked with all the other data on the system – was already available within EMLO. To create the separate space, developer Mat Wilcoxson added the capacity to browse the union catalogue by the gender of senders, recipients, and people mentioned and to filter searches of the entire catalogue by gender in combination with all the other fields of the search interface. Separate catalogue pages were created for WEMLO as a whole and for five new catalogues, totalling over 500 letters to and from English women in the decades around

⁶⁸ A first instalment was announced in the CofK blog for 4 Sept. 2015. Further material is in preparation.

⁶⁹ Project website: <http://brienne.org/>. Partner website: <http://letterlocking.org/> (last accessed 25 June 2019). CofK blog, 10 Nov. 2017.

⁷⁰ James Daybell, *The Material letter in early modern England: manuscript letters and the culture and practices of letter writing, 1512–1635* (London, 2012); James Daybell and Andrew Gordon, *Women and epistolary agency in early modern culture 1450–1690* (Abingdon, 2016).

1600, which Kim prepared from James's data.⁷¹ Arno Bosse also created a separate 'Networks & Resources' hub for WEMLO offering a future exhibition and networking space.

⁷² The launch of WEMLO in the History Faculty on 6 July 2016 was attended by over fifty people from England, Ireland, Scotland, the Netherlands, Brazil, Germany, Sweden, and the US.⁷³ The following year, the Society for the Study of Early Modern Women awarded WEMLO their commendation for the 'Best Digital Scholarship, New Media, and/or Art' created in 2016.⁷⁴

Situating WEMLO within EMLO offered three other advantages. First, it removed the burden of designing, building, and sustaining WEMLO as a self-standing system. Second, it ensured that WEMLO would automatically benefit from any future technical improvements to EMLO. Third, it guaranteed that the ongoing publication of women's letters in EMLO automatically enlarged WEMLO as well, whether or not these catalogues were primarily focused on women. Even before the WEMLO launch, EMLO had attracted a number of significant inventories of women's correspondence, including Antoinette Bourignon, Françoise de Graffigny,⁷⁵ and the Bodleian Student Editions.⁷⁶ After the launch, this pattern continued, often through the involvement of seminar or conference speakers. A calendar of the redoubtable Bess of Hardwick (c. 1521/2 or 1527–1608) was published in December with links back to the deluxe digital edition of her full correspondence created by Alison Wiggins and her AHRC-funded team at the University of Glasgow.⁷⁷ The publication of the catalogue of the first volume of Nadine Akkerman's edition of the letters of Elizabeth Stuart (1596–1662), Queen of Bohemia, was timed to coincide with a workshop on this Scottish princess in St Andrews in April 2017.⁷⁸ The largest increment of women's letters arrived in September 2016 with the publication in W/EMLO of a vast archive of letters to and from the

⁷¹ WEMLO catalogue page: http://emlo-portal.bodleian.ox.ac.uk/collections/?page_id=2595 (last accessed 25 June 2019). Five catalogues: Anne Bacon (c.1528–1610), Margaret Clifford (1560–1616), Anne Dudley (1548/9–1604), Penelope Rich (1563–1607), and Arbella Stuart (1575–1615).

⁷² WEMLO Networks & Resources page: <http://emlo-portal.bodleian.ox.ac.uk/wemlo/>. Facebook page: <https://www.facebook.com/WEMLO-Womens-Early-Modern-Letters-Online-325357000924586/> (last accessed 25 June 2019).

⁷³ On the launch, cf. <https://www.history.ox.ac.uk/article/launch-women%E2%80%99s-early-modern-letters-online-wemlo> (last accessed 25 June 2019) and CofK blog, 6 July 2016.

⁷⁴ On the award, cf. CofK blog, 25 Oct. 2017, and <http://ssemw.org/2017-award-winners/> (last accessed 25 June 2019): 'This impressive website has thoroughly thought through how to digitize women's letters and ensure that they are available to scholars of early modern gender but also to the larger field of early modern scholars.' The result is an open access resource, 'available to scholars at all institutions as well as independent scholars', which helped to ensure that scholarship on women and gender would not be siloed.'

⁷⁵ *Correspondance de Madame de Graffigny*, ed. J. A. Dainard *et al.* (Oxford, 1985-): CofK blog, 15 June 2015.

⁷⁶ Sarah Champone (1 March 2016) before the WEMLO launch; Elizabeth Wagstaffe (12 Dec. 2016) and Elizabeth Compton (30 July 2017) afterward.

⁷⁷ *Bess of Hardwick's Letters: The Complete Correspondence, c.1550-1608*, ed. by Alison Wiggins *et al.*, <http://www.bessofhardwick.org>. (last accessed 25 June 2019). CofK blog, 30 Dec. 2016.

⁷⁸ *The Correspondence of Elizabeth Stuart, Queen of Bohemia*, ed. Nadine Akkerman (Oxford, 2011-). CofK blog, 30 April 2017.

wives of six seventeenth-century Dutch stadtholders.⁷⁹ At the time of writing (June 2019), there are 15,746 letters to or from women in WEMLO, over ten percent of the total of 145,241 letters currently in EMLO.

8. Prospects

This account has focused on CofK's development from a pilot project designed, built, and populated in Oxford to a re-engineered system populated increasingly by a broader scholarly community. But this narrative is incomplete, particularly as an account of CofK's development during Phases III and IV. During the past four years, CofK has moved beyond collaborative population to collaborative systems design and development. Doing so has required revisiting every one of the basic headings discussed above, including the basic EMLO data model, the systems designed to handle it, the community engaged in conceptualising new work in this field, and the resources they are creating. Although space constraints preclude a full account of these more recent developments, a brief summary is needed to understand where the project is headed.

8.1 Collaborative systems design: COST Action IS 1310

The original EMLO data model had been scrutinised by an interdisciplinary team in Oxford and subjected to international inspection from the outset. In order to canvas feedback on a far larger scale, a more comprehensive solution was needed. The solution was found in the COST Action IS 1310, chaired by Howard Hotson and Thomas Wallnig (Vienna), entitled 'Reassembling the Republic of Letters, 1500-1800: A Digital Framework for Multilateral Collaboration on Europe's Intellectual History'.⁸⁰ COST (Cooperation on Science and Technology) – a spin-off of the European Science Foundation subsequently incorporated into Horizon 2020 – does not fund research, resource creation, or systems development. Instead, it funds the large-scale networking activities needed to ensure that major research projects independently funded in different EU countries coordinate their activities in mutually beneficial ways. This was a funding scheme ideally structured to assemble a community of nearly 200 librarians, archivists, scholars, and IT experts from 33 European countries for four years of structured discussion of all the issues involved in rolling out something like EMLO on an international basis.

⁷⁹ This project represents another major collaboration between CofK and the Huygens ING: the 3,959 letter records on EMLO, compiled by a team under the direction of Ineke Huysman, link back to high-resolution scans on the Huygens's servers in Amsterdam. An accompanying virtual exhibition provides portraits and biographical information for each of these six women, a selection of their letters, and illustrations of the palaces in which they lived, worked, and raised their families. Catalogue pages: http://emlo-portal.bodleian.ox.ac.uk/collections/?page_id=2756. Virtual exhibition: <http://emlo-portal.bodleian.ox.ac.uk/exhibition/sw/> (last accessed 25 June 2019). Review of the exhibition: Rebecca Tucker, 'The Wives of the Stadtholders', *Early Modern Women* 12.2 (2018), p.171-4; Reception: CofK blog, 2 Sept. 2016.

⁸⁰ Action website: <http://www.republicofletters.net/> (last accessed 25 June 2019). Dobrochna Futro deserves special commendation for her superb work in simultaneously administrating COST and CofK during Phase III and part of Phase IV, under the unstinting oversight of vice-chair Thomas Wallnig, whose initiative, dedication, and splendid capacity for organisation were fundamental to the success of this demanding project.

In concrete terms, the result was a series of three conferences, four training schools and design sprints, 16 working group meetings, and over 40 short-term international exchanges for early career researchers. From one perspective, this series of events between April 2014 and April 2018 can be regarded as the continuation of the original series of Mellon-funded seminars, workshops, and conferences from 2009 to 2013. Unlike the earlier series, however, the COST-funded events moved beyond Oxford to Brussels, Budapest, Chester, Como, Dublin, The Hague, Lancaster, Lisbon, Oxford, Prague, Tallinn, Valletta, Vienna, Warsaw, Wolfenbüttel, and Zagreb.

As well as taking community-building to the next level, the core task of this project was a large-scale exercise in collaborative systems design. Discussions focused primarily on the standards needed for the temporal, spatial, personal, social, and material aspects of correspondence networks; the systems needed to assemble, reconcile, and exchange data; and the kinds of scholarship which might emerge from a new generation of digital infrastructure shared by a potentially global community of scholars, repositories, systems developers, and designers. The recommendations emerging from these discussions have been outlined in detail a collaboratively written book, which is intended to help provide a framework for the next phase of international collaboration in this field.⁸¹

8.2 Collaborative systems development: Early Modern People, Places, and Dates

By the time the COST Action began in April 2014, CofK had already embarked on collaborative systems development. The first main step in this direction was a partnership with Aalto University (Helsinki) in the development of an open-source tool (subsequently dubbed ‘Recon’) for the semi-automatic reconciliation of people, places, and works in catalogue metadata.⁸² Thanks to the international and interdisciplinary discussions emerging from the COST Action, EMLO is slowly evolving from a monolithic system designed, built, and populated in Oxford, into a component of a distributed, transnational digital infrastructure designed, built, and populated in concert with national and international partners. A second step in this direction is the close partnership with the Huygens ING designed to develop EM People, EM Places, and EM Dates as the basis of a humanities-focused, Linked Open Data-centred infrastructure.⁸³

In Phase IV (2017-19), CofK took the bold decision to disaggregate the original EMLO letter model. Originally, EMLO had regarded letter records as its primary concern, and person and place records as integral components of letter records. Yet a letter record is not elemental: accurate letter metadata rely on the identification of people, places, and dates; and each of these identifications raises difficulties that must be confronted if large-scale digital infrastructure is to be firmly founded. Dating early modern letters accurately requires systems for navigating a complex landscape in which different places transition between different calendars at different times. Locating letters on that landscape requires the capacity to capture the ways in which place names and their situation within larger geographical entities change over time. Disambiguating and linking letter writers and recipients in multiple data

⁸¹ Howard Hotson and Thomas Wallnig, eds., *Reassembling the Republic of Letters in the digital age: standards, systems, scholarship* (Göttingen, 2019): open access version available at <https://doi.org/10.17875/gup2019-1146>.

⁸² Cf. <https://github.com/jiemakel/recon> (last accessed 25 June 2019) and the discussion in Eero Hyvönen et al., ‘Reconciling Metadata’, in *Reassembling the Republic of Letters*.

⁸³ <https://timbuctoo.huygens.knaw.nl> (last accessed 25 June 2019).

sets requires the development of authority files for huge numbers of individuals who are not found in national biographical dictionaries and of prosopographical structures for collecting precious scraps of information on them.

With these challenges in mind, CofK's pursued a six-fold agenda during Phase IV: (1) to decouple our systems for handling early modern people, places, and dates from the core system for handling letter records; (2) to define new capabilities for these new systems through dialogue with a wide range of partners; (3) to launch Early Modern Dates as a free-standing web-service able to reconcile multiple calendars; (4) to prepare Early Modern Places as a free-standing, Linked Open Data resource independent of EMLO; (5) to continue work on a data model for Early Modern People; and (6) to scope out how to integrate these databases with EMLO and other relevant Linked Open Data resources.

In the course of Phase IV, the vastly different scales of these challenges became evident. The core challenge of modelling dates is mathematical and therefore ultimately solvable. EM Dates will therefore be released at the end of Phase IV as a web-service and API for converting between the range of early modern calendars. Modelling places is more complex, and EM Places will only be brought to an advanced stage of planning as a Pelagios-compatible, historical gazetteer with support for temporal places, drawing on the c. 6,000 disambiguated early modern place records recorded in EMLO. In addition to providing a firmer grounding for EMLO, these resources are designed to serve the needs of early modern studies more generally.⁸⁴

8.3 Collaborative data model design and development: towards a prosopographical data model

Modelling early modern people is more complicated still and therefore requires separate discussion. There is little point in assembling a catalogue of the letters without identifying the letter writers and recipients. Consequently, most major projects in the field collect biographical records alongside epistolary ones. But early modern people are difficult entities to model, and are poorly served by existing authority files. Correspondence networks are usually scale-free distributions, in which very large numbers of letters are exchanged with a very small number of correspondents, and very small numbers of letters are exchanged with very large numbers of correspondents, the vast majority of whom are missing both from standard authority files (such as VIAF and CERL) and from the national biographical dictionaries and bibliographies on which they principally draw.

This situation produces three basic requirements which a system for handling people in the republic of letters must meet. First, in order to link records to an individual found in multiple correspondences, an enhanced authority file is needed which could be populated, in the first instance, with the tens of thousands of unique early modern person records already collected in EMLO, the Electronic Enlightenment, and cognate projects. Accurate linking, however, requires the capacity to disambiguate people with similar names and titles. Such disambiguation can only be undertaken with reference to the additional biographical details needed to distinguish people from one another. So the first requirement gives rise to the second: namely, the need for a data model with which to assemble scattered scraps of biographical data on obscure correspondents for easy interrogation by computers as well as scholars. Much of this data will record non-epistolary encounters between citizens of the republic of letters in the course of their education, travel, careers, and social lives; so this second requirement naturally expands into a third: namely, the need for a data model capable

⁸⁴ These projects are described in detail in *Reassembling the Republic of Letters*, II.2-3.

of capturing data on all the different kinds of relations between individuals in a given group, epistolary and otherwise.

Incremental work on a solution to these problems has been ongoing within the CofK community for many years. Howard Hotson began collecting scraps of data on Hartlib's correspondence long before the advent of CofK. In Phase I, Leigh Penman enhanced this material and developed some of it into more mature biographical entries. In Phase II, Tanya Grey Jones helped develop an event-based prosopographical data model in RDF which Iva Lelkova and Robin Buning began populating in spreadsheet form. In Phase III, Martin Hadley piloted interactive visualisations of this data.⁸⁵ During the COST Action, Jetzte Touber and Mikkel Munthe Jensen critically reviewed the CofK prosopographical data model in one pair of STSMs, Dagmar Mrozik and Gabriella Martinez developed complementary material in a second pair, and the design sprints in Como experimented with solutions to some aspects of these problems.⁸⁶ In July 2018, the third 'Digitising Enlightenment' colloquium met in Oxford to focus especially on the problem of prosopography;⁸⁷ and in May 2019 CofK co-sponsored with the British Academy a second colloquium on prosopography in University College London.⁸⁸ Some basic proposals, emerging from this work, are outlined in two chapters of *Reassembling the Republic of Letters*; but a great deal of collective effort will be needed to generate a robust and consensual solution to this problem.⁸⁹

8.4 Collaborative population of EMLO: COST STSMs and associated grant applications

The efforts in community building integral to the COST Action also advanced the collaborative population of EMLO. The most direct channel was via the COST-funded 'Short-Term Scientific Missions', which funded collaborative visits by emerging scholars to partners abroad. The missions facilitated development of Robin Buning's inventory of the correspondence of Isaac Vossius, Vittoria Feola's inventory and digital archive of the Bartolomeo Gamba collection, and Zeljka Salopez's preparation of metadata from a digital archive of Croatian neo-Latin correspondence. Other missions have provided valuable surveys of rich resources still poorly represented in EMLO, such as Justine Walden's wide-ranging census of letter collections in Italy.⁹⁰

The culmination of these efforts were five STSMs between February 2016 and March 2017 which allowed emerging scholars from Utrecht to visit Oxford for consultation and training in assembling and analysing data on early modern printed letter collections.⁹¹ This

⁸⁵ <https://idn.web.ox.ac.uk/article/cultures-knowledge-case-study> (last accessed 25 June 2019).

⁸⁶ Reports on the first three of these STSMs can all be found at <http://www.republicofletters.net/index.php/stsm/> (last accessed 25 June 2019). See also the contributions by Vladimír Urbánek, Jensen, Mrozik, and Martínez in Hotson and Wallnig, eds., *Reassembling the Republic of Letters*, ch IV.3 section 3, and ch. IV.4, sections 2.2, 3.2, and 4.2.

⁸⁷ <https://voltairefoundation.wordpress.com/2018/09/25/digitizing-enlightenment-iii/> (last accessed 25 June 2019).

⁸⁸ <https://tinyurl.com/y4o9srhc> (last accessed 25 June 2019).

⁸⁹ Hotson and Wallnig, eds., *Reassembling the Republic of Letters*, chs. II.4 and IV.4.

⁹⁰ A hyperlinked list of STSM reports can be found at <http://www.republicofletters.net/index.php/stsm/>. The Croatian project is 'Croatiae auctores Latini': see <http://www.ffzg.unizg.hr/klafil/croala/> (last accessed 25 June 2019).

⁹¹ For the link with COST, see the STSM reports of Lara Bergers, Emma Mojet, Roberto Bellignacci, Mandi Astola, and Celine Frohn between Feb 2016 and March 2017, listed at

work helped lay the foundations for one of the most ambitious and important new initiatives in the field: Dirk van Miert's five-year, ERC-funded project, 'Sharing Knowledge in Learned and Literary Networks (SKILLNET): The Republic of Letters as a Pan-European Knowledge Society' (July 2017 to June 2022). One of SKILLNET's fundamental tasks is to develop tools and workflows for crowd-sourcing large quantities of metadata on the letters in these printed epistolaries for eventual publication in EMLO.⁹²

SKILLNET caps a list of successful independent grant applications devoted in part to populating EMLO. These include the inventory of Johann Valentin Andreae's correspondence funded by the Deutsche Forschungsgemeinschaft in Wolfenbüttel; Jackie Stedall (†) and Philip Beeley's three-year AHRC grant for a study of 'The Life and Letters of John Collins' in Oxford;⁹³ Paul Botley's four-year Leverhulme-funded project in Warwick which will edit 731 letters written between 1610 and 1614 to and from Isaac Casaubon;⁹⁴ and Bernhard Schirg's project, funded by the Volkswagen-Stiftung at the Forschungszentrum Gotha der Universität Erfurt, on the Swedish polyhistor, Olaus Rudbeck (1630–1702).⁹⁵ Another promising development is the three-year doctoral fellowship, funded by the Digital Humanities team at the Institut des sciences du calcul et des données (ISCD) of the Sorbonne, on correspondence networks relating to the history of science and mathematics.⁹⁶

8.5 Collaborative network analysis

Many of these strands are woven together in an AHRC-funded project entitled 'Networking Archives: Assembling and analysing a meta-archive of correspondence, 1509-1714'.⁹⁷ Between October 2018 and June 2021, this project will take these partnership arrangements to the next level by uniting the core CofK team with a highly successful project already funded by the AHRC: namely, the 'Tudor Networks of Power' project (TNofP) led by Ruth Ahnert (Queen Mary, University of London) with Sebastian Ahnert (Cambridge) and designer Kim Albrecht (Harvard and Potsdam).⁹⁸ Data creation will again be fundamental, but pursued on a larger scale and at a faster pace. To the 140,000 records currently on EMLO will be added two roughly commensurate catalogues: records of the c. 133,000 letters in the central archive for English history in the Tudor period, the English Domestic and Foreign State Papers for 1509-1603, already curated by TNofP; and records of c. 200,000 letters from the analogous archive for the Stuart period (1603-1714), which will be curated afresh. Collaborative systems development will help speed the curation, disambiguation, and

<http://www.republicofletters.net/index.php/stsm/>. For its later place within SKILLNET, see <https://skillnet.nl/crowdsourcing-erol/> (last accessed 25 June 2019).

⁹² SKILLNET website: <https://skillnet.nl/> (last accessed 25 June 2019). CofK blogs: Sept. and Oct. 2017. The first fruits of this reciprocal collaboration have already appeared: see note 62 above

⁹³ See <http://www.culturesofknowledge.org/?p=3964> (last accessed 25 June 2019).

⁹⁴ Cf. https://warwick.ac.uk/fac/arts/ren/projects/casaubon/the_project/ (last accessed 25 June 2019) and the CofK blog: 5 March 2014.

⁹⁵ Entitled 'Reaching for Atlantis. The cultural biographies of objects under the Swedish Empire and beyond'. CofK blog: 30 Aug. 2017. Preliminary website: <https://www.reachingforatlantis.de/index.php/beispiel-seite/> (last accessed 25 June 2019).

⁹⁶ CofK blog: July 2018.

⁹⁷ See <https://networkingarchives.org> (last accessed 25 June 2019).

⁹⁸ Grant period: Jan. 2015 – Dec. 2017: <https://ahrc.ukri.org/research/case-study-archives/tudor-networks-of-power-1509-1603/> (last accessed 25 June 2019).

reconciliation of these huge quantities of data. This will create an open-access meta-archive of nearly half a million records, the largest consolidated set of curated epistolary records available anywhere for the early modern period.

The tools and algorithms previously developed by TNoP for quantitative network analysis will then be combined with traditional approaches to study the kinds of networks formed when multiple scholarly correspondences intersect with one another and with a massive archive like the State Papers. These tools will be enhanced with a new set of advanced algorithms and scripts designed to examine the evolution of these networks over time. Thematically, the project will focus on ‘intelligencing’ activity of figures like Samuel Hartlib, whose activities took place at the intersection between the international, scholarly republic of letters (documented in EMLO) and the English commonwealth and republic of the 1640s and 50s (documented in the English State Papers). Building on the experience of the COST Action, the mentoring of emerging scholars will continue with a series of training schools, a colloquium, and an edited volume. These training schools will develop a curriculum to support the teaching of network analysis. Individual components of the project's infrastructure (data, software, documentation, and methods) will be consolidated and shared for reuse. Last but not least, the project is designed to pilot a means of reversing the steady proliferation of stand-alone digital ‘data silos’ by demonstrating how merging successful initiatives and pooling data and tools can help create sustainable infrastructure for interdisciplinary research.

9. Conclusion

Cultures of Knowledge began as a primarily editorial project with a more speculative digital coda. Even when emphasis shifted primarily to the digital side, EMLO was not merely a technical experiment or an exercise in resource creation. Its key innovation was rather to attempt to assemble scholarly resources in a manner midway between a large-scale, stand-alone, collaborative research project and full-scale crowd-sourcing. It is therefore best described, in retrospect, as a social and cultural experiment, undertaken to explore the conditions under which members of a scholarly community will voluntarily contribute their findings to a new kind of shared resource. The hypothesis it set out to test was not a thesis about the early modern republic of letters per se but an intuition about the contemporary scholarly community: namely, that technology could help create the conditions in which scholars, projects, publishers, and potentially also libraries and archives would pool their correspondence data, reverse the proliferation of data silos, and gradually create the big data on which the next generation of tools, methods, and scholarly insights will be based. The title ‘Cultures of Knowledge’, first suggested by Pietro Corsi, was therefore strangely prescient; but it ended up applying, less to the task of distinguishing a variety of cultures of knowing within the seventeenth century, than to creating an environment within which new cultures of assembling, exchanging, pooling, curating, publishing, and analysing knowledge resources can be nurtured along with the new forms of knowledge arising from them.

What then is the result of this experiment? EMLO is well on its way to becoming a meta-archive, bringing together many archives; and the potential for its further expansion is in principle almost unlimited. Even at this early stage of transition, the quantity of metadata available is enormous. Most publishers, and many projects and repositories, are keen to contribute data to an openly accessible meta-archive, and this result provides proof of concept for further phases of more ambitious experimentation. But two major obstacles remain. One is the hesitation of certain potential partners to release data needed for any genuinely

representative dataset of more than discrete portions of the republic of letters as a whole. This hesitation is often a consequence of their institutional responsibility for preserving data on some specific aspect of our broader cultural heritage. A second, related obstacle is that no one project has the resources – intellectual or financial – to assemble, curate, preserve, and publish more than a fraction of the relevant data. The ultimate solution to the problem tackled by EMLO is therefore the creation, not of a single, enormous, monolithic data store, but of a new kind of distributed infrastructure within which scholars can access and interact with data assembled from a digitally interlinked network of data repositories which each have the responsibility and resources for curating parts of a larger whole. For CofK and the COST Action arising from it, the ultimate success would be to help create the conditions for the development of such distributed infrastructure, even if doing so rendered many features of EMLO obsolete.

Preparing such condition demands more than mere technology: it also requires returning in conclusion to the theme of cultures of knowledge in transition. In a lecture on ‘Digital Transformations’ delivered in Oxford a few years ago, Andrew Prescott argued that the digital transformation of humanistic research and teaching was something still to be awaited in the not too distant future.⁹⁹ The reason is that revolutionary technology is not sufficient to precipitate revolutionary change: it must be applied in a manner capable of bringing a more gradual *cultural* change in its wake. While technical change, per se, is often very rapid, relatively easy to engineer, and even surprisingly inexpensive, cultural change is slower and more difficult to manage. In order to work out how best to use technology, individual research communities must undertake collaborative experiments and be prepared to undergo culture shifts.

Moreover, the culture must shift at several different levels: individual, institutional, and systemic. Facilitating changes in individual behaviour is relatively simple: one must heighten the advantages of changed behaviour and lower the obstacles to change. This is where the work of CofK to date has primarily been focussed. The task of lowering the barriers to participation is well underway. When working within the EMLO environment, scholars no longer need to devise a data model, to construct spreadsheets or databases, to create all the people or place records they need *de novo*, or even to hunt manually through standard authority files for links with which to create new people and place records: instead, a range of possible matches is provided automatically by a tool designed for this purpose. With further development, they may also not need to answer every single scholarly question unassisted: using a forum system, it will be possible to raise questions of detail with the entire scholarly community working within the resource. As more data is entered into the system, there is also less need to start the hunt for correspondence from scratch, since the nuclei of individual correspondence catalogues has begun to accumulate spontaneously. There is no need to design, fund, or build online systems for displaying, publishing, searching, and sorting the data, or to invest further time and money in attempting to attract attention to new web resources: the existing channels and networks surrounding the project are designed to serve these functions for all contributors. Best of all, there is no need for a host of small projects to fund periodic systems up-grades as technology changes, or to make their data and systems sustainable in perpetuity. Whatever the long-term solution to the sustainability problem, it will have to involve replacing a constantly proliferating galaxy of often poorly configured bespoke resources with a more compact network of large, collaboratively funded and populated resources networked to serve as national and international infrastructures.

⁹⁹ Andrew Prescott, ‘Digital Transformations’, TORCH Lecture, St. Anne’s College, Tuesday, 17 Feb. 2015.

Technical facilities of this kind, if carefully crafted to meet the needs of scholarly communities, are capable of changing the publishing preferences of individual scholars. But individual scholars cannot change their behaviours without a shift in the cultures of the institutions in which they operate; and at this point we encounter the second-level difficulty of overcoming institutional inertia. For five centuries, modern academic culture has been dominated by print. Within the humanities, print publications retain an authority which digital publications still lack. Moreover, the two main formats of print publications – book-length monographs and much shorter journal articles – exercise a virtually uncontested duopoly over what counts as bona fide scholarly productions. Publications larger and different in character than the monograph, smaller than the article, and indeed midway in size between these two, are difficult to publish and are not accorded the same regard; and these problems are compounded by the lingering gap in prestige between digital publications and the best established print alternatives. At the same time, sole authorship is assumed and rewarded, and richer forms of collaboration are either silently appropriated by a lead author or devalued in proportion to the number of collaborators.

These problems of cultural transition are further exacerbated, at a still higher systemic level, by the two very different ways in which academic production management regimes (like the Research Assessment Exercise and the Research Excellence Framework in the UK). While monographs and articles are assessed primarily as contributions to academic scholarship, large collaborative undertakings are assessed according to their measurable ‘impact’ on public life outside academia altogether. Missing from this regime is a means of assessing the purely academic value of large-scale scholarly undertakings, above and beyond their ‘impact’ on the public sphere. Without such a measure, funding of the core purpose of curating, publishing, and preserving epistolary data and metadata will remain in short supply. At the same time, the emphasis on institutional competition deliberately fostered by these regimes undercuts the impulse toward extramural collaboration which is the precondition for any really satisfactory solution to large-scale scholarly problems.

Given the need for change at several different levels, the cultural change integral to ‘digital transformation’ will not be easy or instantaneous. As the experience narrated above shows, creating community-curated digital resources requires time: time to develop and test models and tools, time to make mistakes and try again, time to allow a community to form organically, time to involve that community in the development of tools and procedures, and above all time to negotiate the terms and conditions which constitute new cultures of knowledge within this particular domain. Given this long timeline, the crucial precondition for the success of an enterprise of this kind is something very rare. What is needed to begin this process is a source of potentially long-term funding, coupled with the flexibility to reallocate and roll over funding as priorities evolve, and a clear and relatively predictable process for reapplication to fund successive phases of work. Cultures of Knowledge is exceptionally fortunate to have found such a source of funding in the Andrew W Mellon Foundation of New York, without which this project could not possibly have reached its current stage.

This good fortune imposes on the project team a sense of obligation to keep the project moving forward toward the objective articulated at the outset: to help create a humanistic research project ‘capable of transforming disciplines’. Looking forward, the single biggest question is sustainability. Collecting hundreds of data silos together greatly simplifies the answer to this question, but everything still depends on rendering sustainable the meta-archive which unites them. Solving this problem may involve higher-level convergence with major projects in a still broader field, starting with some of the other projects included in this

volume. The series of colloquia on Digitising Enlightenment has already played an important role in this process.