



## Briefing document on COST Action IS1310

I. General Introduction.....	1
II. Introduction to Working Groups.....	3
III. Working Group Subheadings: Overview.....	4
IV. Working Group Abstracts and Agendas.....	5

# I. General Introduction

---

## GENERAL OBJECTIVES

Between 1500 and 1800, the evolution of postal communication allowed ordinary men and women to scatter letters across and beyond Europe. This exchange helped knit together what contemporaries called the *respublica litteraria*, a knowledge-based civil society, crucial to that era's intellectual breakthroughs, and formative of many modern European values and institutions.

Ironically, the exchange of letters which created this community also dispersed the documentation required to study it, posing enormous difficulties for historians of the subject ever since. To reassemble that scattered material and chart the history of that imagined community we need a revolution in digital communications.

This COST Action is dedicated to envisaging the open-access, open-source, transnational digital infrastructure capable of facilitating the radically multilateral collaboration needed to reassemble this scattered documentation and to support a new generation of scholarly methods and research questions.

## ABOUT

**COST** is an intergovernmental framework for [European Cooperation in Science and Technology](#). It exists to coordinate on a European level ongoing research that is funded nationally. Rather than funding research itself, resource creation, or IT systems development, COST Actions provide the networking support needed to ensure that nationally funded initiatives add up to something greater than the sum of their individual parts.

**COST Action IS1310** *Reassembling the Republic of Letters, 1500–1800* is essentially a networking programme therefore. It emerged in response to the expectation that that the ongoing revolution in digital communications technology can solve the scholarly problem created by the evolution of postal communication in the early modern period: the problem, namely, of piecing back together corpora of manuscript correspondence deliberately scattered across and beyond entire continents. In essence, a new scholarly network is needed in order to study the older scholarly network of the early modern period. To sustain this new network, we need a new breed of digital networking tools; and the purpose of this Action is to assemble the network needed to design these tools: a network to devise a networking platform to support a network studying past networks.

**The objectives** of this Action are, therefore, essentially two-fold: technical and historiographical. The technical objective is to plan a state-of-the-art digital system within which to collect a pan-European pool of highly granular data on the Republic of Letters. This involves designing tools for collecting, standardizing, navigating, analysing, and visualizing unprecedented quantities of epistolary data, and for facilitating new forms of international and interdisciplinary scholarly collaboration, thereby consolidating a new virtual Republic of Letters. Neither technical innovation nor resource creation, however, are undertaken as ends in themselves: rather, they are devoted to serving the second, historiographical agenda, which both generates the fresh research questions needed to design the infrastructure and uses the emerging technology to devise new

methods, pose new questions, and answer old ones.

**The scope** of an Action capable of pursuing these objectives is therefore necessarily both pan-European and highly interdisciplinary. As of March 2015, [30 of the 35 COST member countries](#) are formal partners to this Action. This community is likewise highly interdisciplinary: as well as scholars from numerous humanistic disciplines, it relies on the combined expertise of archivists, librarians, and specialists in a wide range of digital technologies as well as visualization, communication, and intellectual property law.

**The work** of the Action is conducted primarily in six Working Groups and a series of annual Conferences. Dividing the technical agenda between them, the Working Groups seek collectively to address the main technical issues involved in devising transnational digital infrastructure to serve this field, as well as the legal agreements and scholarly conventions which that infrastructure requires. Agreeing this agenda and distributing the work involved in pursuing it is the agenda of the Action's first Conference, which meets in Oxford in March 2015. The parallel historiographical agenda will be the focus of the second conference, in 2016. These two agendas must proceed in parallel, since the relationship between them is reciprocal: just as the infrastructure must be designed to address scholarly research questions, possibilities opened up by the new infrastructure must shape the historiographical agenda in turn.

**The activities** funded by COST also include Workshops held at irregular intervals depending on need, Training Schools designed to induct younger scholars into the techniques and methods being pioneered by the group, and Short Term Scientific Missions (STSMs), which fund international exchange visits within the network likewise targeted at colleagues in early career. Invitations to apply for Training Schools and STSM will be posted on the News section of the website and circulated to all those on the Action's email list.

**Participation** in the Action can take a number of different forms. The Action is coordinated by a Chair and Vice-Chair and administered by a Grant Manager. Responsibility for the Working Groups is devolved to six WG Leaders, who also make up the Action's Steering Group, together with the Chair, Vice-Chair, STSM Coordinator, Webmaster, and the Coordinator of the forthcoming conference. Ultimate decision-making authority rests with a Management Committee, composed of up to two members from each participating country, each of whom can nominate one substitute to represent them when they are unable to attend. Every Management Committee Member is also a member of a Working Group, which also include a small number of additional Working Group Members specially recruited to add necessary expertise. Affiliates to the Action can also participate in discussions via the Members Only portion of this website. Scholars in early career can also participate via Training Schools and STSMs. The Contacts page allows interested parties to join the Action's email list and receive a regular newsletter and other bulletins.

## II. Introduction to Working Groups

---

At the centre of this Action are six Working Groups, the scope of which is determined by the basic features of letters themselves. Letters are *written documents* sent by a one or more *person(s)* at a specific *time* and *place* to one or more other *person(s)* in another *place*. The spatial and temporal, personal, textual, and material features of letters structure the agenda of four Working Groups; while the other two Working Groups are devoted to the means of bringing data on all these dimensions together and rendering it visible and intelligible to a variety of audiences.

### Working Group 1 **Space and Time**

WG 1 considers how to identify, represent, analyse, and visualize the data on the spatial and chronological dimensions of correspondence networks found in catalogue records, in letter texts, and in prosopographical data on letter writers and recipients.

### Working Group 2 **People and Networks**

WG 2 studies how best to structure and assemble biographical data describing the citizens of the republic of letters, as well as how to structure, analyse, and visualize data on the networks created by epistolary and related forms of learned exchange.

### Working Group 3 **Texts and Topics**

WG 3 focuses on the presentation of letters as images and digital texts, the development of tools for transcribing, annotating, editing, text mining, and topic modelling them, the translation from print to digital form and vice versa, and the use of training schools to help specify a virtual research environment adapted for the study of this material.

### Working Group 4 **Documents and Collections**

WG 4 will develop means of describing the physical characteristics of letters, of capturing information on their provenance history, and of assembling information on collections of learned correspondence and existing aids for finding them.

### Working Group 5 **Data Exchange and Strategic Planning**

WG 5 deals with the means of generating standardized digital data of all these kinds in unprecedented quantities, the technical and legal arrangements necessary for exchanging them between participating individuals, projects, and institutions, and the problems of funding and long-term preservation and sustainability.

### Working Group 6 **Visualization and Communication**

WG 6 will conceptualize and specify a new generation of visualization tools applicable to all stages in the process of working with epistolary data, and will consider means of communicating both the scholarly and the technical interest of the Action to a variety of different audiences, within and especially beyond the academy.

Each Working Group breaks down these large issues into component problems and pursues consensual solutions with them in discussion with archivists, librarians, IT experts, and scholars from a variety of fields drawn from across and beyond Europe.

### III. Working Group Subheadings: Overview

---

#### **WG 1: Space and Time**

- I. Cognate projects and research questions
- II. Metadata standards and authorities
- III. Data and metadata within text corpora
- IV. Queries
- V. Mapping and other visualization options
- VI. Pilot Projects

#### **WG 2: People and Networks**

- I. Cognate projects and research questions
- II. Personal name authorities
- III. Proposographical data model and input form
- IV. Automating provision of basic prosopographical data and metadata standardization
- V. Network analysis
- VI. Visualization tools and pilot projects

#### **WG 3: Texts and Topics**

- I. Fundamentals: Images of documents, text coding
- II. Transcription, Annotation, and Editing
- III. Textual analytics: data mining and topic modelling
- IV. Training Schools: testing tools, training scholars
- V. Toward a Virtual Research Environment

#### **WG 4: Documents and Collections**

- I. Refining the data model
- II. Census of correspondence collections: in manuscript and print
- III. Crowd-sourcing catalogue level descriptions

#### **WG 5: Data Exchange and Strategic Planning**

- I. Generating digital metadata: crowd-sourcing, scanning and keying
- II. Uniting metadata standards
- III. Digitizing learned correspondence: scoping a master plan
- IV. Sharing digital metadata: legal agreements and scholarly conventions
- V. Recruiting contributions
- VI. Preservation and sustainability: technical and financial challenges

#### **WG 6: Visualization and Communication**

- I. Existing tools and approaches: an overview of previous work
- II. Conceptualizing and specifying new tools and approaches
- III. Communication within the scholarly community: website, networking site, journal
- IV. Communication with broader publics

## IV. Working Group Abstracts and Agendas

---

### WG1: Space and time

#### ABSTRACT

Digital technology provides unprecedented opportunities for understanding the spatial and temporal dimensions of exchanges of learned correspondence, of the movements of correspondents themselves, and of the discussions taking place within the letters they exchange. The most basic precondition for realizing this potential is the development of standards for presenting spatial and temporal information in catalogue records (II) and in letter texts themselves (III). This will facilitate the development of techniques for analysing large collections of epistolary, prosopographical, and textual data and metadata (IV) and visualizing the results (V), potentially in one or more of pilot projects (VI). At the heart of WG 1 will be, therefore, a dialogue between historians, geographers, and IT experts aimed at identifying ways in which fresh scholarly questions (I) can be answered through the application of geographical information systems (GIS), network analysis (in partnership with WG 2), natural language processing (in partnership with WG 3), and visualization strategies (in partnership with WG 6).

WG 1 is led by [Ian Gregory](#), Professor of Digital Humanities in the Department of History at Lancaster University. A geographer by training specializing in the application of geographical information systems to humanistic materials, he holds a Starting Researcher Grant from the European Research Council in support of a project entitled '[Spatial Humanities: Texts, GIS, Places](#)'.

#### AGENDA

##### I. Research questions and cognate projects

1. Collect and study a *gallery of projects* applying digital technology to mapping and analysing aspects of 'intellectual geography'.
2. Use this gallery to help scholars, editors, and others *identify the questions* they want digital technologies to help them answer. This process is fundamental if this WG is to make a real difference to the study of early modern Europe, and will therefore be iterated at appropriate stages during the Action.
3. Assemble and prioritize a *list of functionality and applications* for dealing with data on the republic of letters.

##### II. Metadata standards and authorities

1. Determine the *structure* of spatial and temporal metadata within catalogue records.

2. Choose *existing gazetteers* as baseline authorities (e.g. [Geonames](#), [Getty Thesaurus of Geographical Names](#), [Orbis Latinus](#)).
3. Develop means of creating *new sub-gazetteers* for extending, enhancing, and correcting existing gazetteers. Current gazetteers are very general and not very well adapted to multi-lingual historical data, let alone the needs to specific corpora. Sub-gazetteers that enhance these for specific countries or collections may be an appropriate solution.
4. Explore options for the *reciprocal exchange of metadata* with gazetteers selected in stage II.2. The [collaboration](#) of the Electronic Enlightenment with the Getty Research Institute provides a valuable precedent.
5. Investigate *semi-automated matching* of name variants and geographic coordinates. Cultures of Knowledge plans to pilot such semi-automated matching from April 2015 onward.

### III. Data and metadata within text corpora

1. Determine how spatial and temporal metadata should be *encoded within texts*.
2. Devise *automated techniques for identifying and encoding* place-names and temporal references within textual corpora, building on the work of the [Spatial Humanities](#) project.
3. Devise means of semi-automating (II.5) the *enrichment of sub-gazetteers* (II.3) with fresh nomenclature extracted from texts.

### IV. Queries

1. Building on the questions generated in item I.2, explore methods of conducting ‘distant’ *analyses of large volumes of structured catalogue metadata*. As well as mapping the correspondence of individuals, this might involve mapping all of the places from or to which letters were sent from a given location during a specific interval, or (to take a more complex example) graphing the times at which letters in a given collection are sent from or near a given place.
2. Investigate queries combining spatio-temporal catalogue metadata with *geographical references identified within letters texts* (in the manner indicated in item III.2). For instance, ‘How does the distribution of places mentioned in a correspondence change over time?’
3. Investigate the potential for *combining geo-spatial data with thematic data* generated by topic modelling in WG 3 — for instance, the way in which the discussion of specific themes travelled across Europe via learned correspondence.

### V. Mapping and other visualization options

1. Building on the gallery assembled at stage 1, consider (in partnership with WG 6) the *means of visualizing geo-spatial data*, whether cartographically or otherwise. The methods, conventions, and technologies already developed in historical geography must be explored to the full in order to build on best practice.
2. Consider the range of *background information* to be provided on map underlays to aid the analysis of spatial and spatio-temporal data and metadata: physical characteristics

(topography, rivers), urbanization, transport links (roads, canals), commercial networks, territorial boundaries, confessional geography, patterns of academic mobility, military movements, and dynastic or diplomatic networks. How might such information be coded and represented so that users can toggle through visualization options in an interactive and user-friendly manner?

3. Address the problem of *visualizing uncertainty and gaps in data*, drawing in particular on the experience of Stanford's [Humanities+Design](#).
4. Explore (in partnership with WG 6) the graphic and technical options for *visualizing temporal sequences* within structured data, including but not restricted to interactive presentations and animation.

## **VI. Pilot Projects**

1. Conduct one or more pilot projects on the above based on suitable (sub) corpora, illustrating potential for larger follow-up projects.



## WG2: People and Networks

### ABSTRACT

The most complicated entities involved in the exchange of learned letters are the people who exchange them and the networks created by the exchange. This complexity is compounded by the fact that people and networks cannot really be separated: typically, the exchange of letters rests directly or indirectly on pre-existing networks of social exchange, so correspondence networks can be fully understood only with reference to data documenting non-epistolary as well as epistolary contact. After identifying research questions and drawing inspiration from previous work (I), WG 2 will devise a data model for the prosopographical information required to answer these questions and an input form as clear, simple, intuitive, and flexible as possible (II). It will then identify major electronic sources of relevant structured data, enable access to them (III), and design and create tools to help scholars reconcile and fill in gaps in their data in a semi-automated fashion (IV). The final task will be to devise tools for visualizing and analysing data sets — from individual people to multi-dimensional networks — in order to answer their research questions (V).

WG 2 will be responsible also for establishing the **technical grounding** of this infrastructure as a whole, which will be based on the following general principles:

1. Use open Linked Data data models and modularize tools behind open APIs where possible to maximize flexibility and reusability;
2. Where possible, build upon open source tools and release own contributions as open source;
3. In the future, following these two considerations will enable
  - a. new actors to join the network more easily,
  - b. new visualization and other tools to be developed on top of the framework with less effort,
  - c. gradual improvement and evolution of the tools and infrastructure itself, and
  - d. gradual improvement of data quality.

WG 2 is led by Eero Hyvönen, Professor in both the Department of Media Technology at Aalto University and the Department of Computer Science at the University of Helsinki. He is also [Research Director](#) of Aalto's [Semantic Computing Research Group](#).

### AGENDA

#### I. Identify research questions and cognate projects (in parallel with WG 1 — Leaders: Ruth Ahnert and Sebastian Ahnert)

1. Review prior work
  - a. Study the harvesting, analysis, and visualization of material from the *Oxford Dictionary of National Biography* undertaken by the '[Six Degrees of Francis Bacon](#)' project. Study the prospects for integrating this approach with the scholarly crowd-sourcing of prosopographical entries. Consider also the feasibility of a similar approach to other leading national biographical dictionaries;

- b. Analysis of *epistolary metadata* alone, which studies the manner in which information can be exchanged through networks documented by correspondence alone (Ruth and Sebastian Ahnert's work with the [State Papers Online](#) for Tudor England);
  - c. Analysis of people mentioned (*co-citation proximity analysis*), which reveals which people were most commonly discussed in proximity to one another in collections of correspondence;
  - d. Analysis of relationships implicit in major biographical sources such as national biographical dictionaries;
  - e. Analysis of relationships within prosopographical data, which maps the data on familial, personal, professional and other relationships captured prosopographically (for instance, in the Cultures of Knowledge project);
  - f. Any combinations of the above.
2. Formulate use cases and research questions (coordinated with item III, identifying sources that could help answer these questions).

## II. Proposographical data model and input form (Leader: Tanya Grey Jones)

1. Review related work on representing people and networks ([VIAF](#), [ULAN](#), [VIVO](#), [BIO](#), [CIDOC CRM](#), [RELATIONSHIP](#)).
2. The basic model is likely to consist of 'event streams', documenting a specific event in the biography of an individual, typically relating to other persons at a specific place and time.
3. The model should focus on the kinds of events typically central to the life of an early modern intellectual, such as schooling, university study, academic travel, membership in learned societies (formal and informal), and stages in learned careers.
4. Reconstructing the social networks underlying correspondence networks likewise requires focus on location and contact histories.

## III. Identify and enable access to electronic sources of relevant structured data (coordinated with WG 1 – Leader: Eetu Mäkelä and CofK)

1. Evaluate known data sources for suitability (coordinated with item I.2)
  - a. Authority files (VIAF, CERL, GND, ULAN, DBPedia, Freebase);
  - b. Publication data (BNF, DNB, BNB, OCLC WorldCat, EEBO, ECCO);
  - c. Appearances in letter metadata (EMLO, [EE](#), CKCC, CEEC);
  - d. Appearances in letter texts (CKCC/ePistolarium, CEEC, EE) [coordinate with WG 3];
  - e. Geographical gazetteers (TGN, Pleiades) [coordinate with WG 1].
2. Enable access to the data sources, either through utilizing existing APIs or importing and creation of such APIs.
3. As a means of reciprocity, arrange equivalent API access to data created as part of the work of WG 2 (e.g. EMLO internal authorities) and negotiate exchange with relevant external agencies (e.g. VIAF/CERL).

## IV. Design and create tools to help scholars reconcile and fill in gaps in their data in semi-automated fashion (Leader: Eetu Mäkelä and CofK)

1. Arrange user tests of EMLO web-form for manual collection of prosopographical data.

2. Investigate semi-automated matching of personal name variants (using e.g. Silk/OpenRefine) on the basis of work undertaken by Cultures of Knowledge from April 2015 onward.
3. Devise automated means of providing editors with *relevant entries in standard biographical dictionaries*, to inform record input, matching and disambiguation. This work will build on the [Letter Metadata Prototype](#) devised in Halle and further work being undertaken by Cultures of Knowledge in 2015.
  - Precondition: compile a *list of web-mounted biographical dictionaries* covering the early modern period (a task for a scholarly sub-committee).
4. Devise automated techniques for identifying and encoding personal names within textual corpora, in order to code letter texts and to enrich letter records with people mentioned (building on work in WG 1 item III.2).

**V. Devise tools for visualizing and analysing prosopographical data** in order to answer their research questions (coordinated by WG 6)

1. Evaluate the suitability of existing tools (e.g. Palladio, Europeana4D, VISU) with regard to answering the research questions derived in item I.2.
2. Develop (open) functionality to fill gaps in existing tools.
3. This exploration will be enriched and informed by conducting one or more pilot projects based on suitable (sub)corpora, illustrating potential for larger follow-up projects, for instance:
  - a. Epistolary metadata + people mentioned: Circulation of Knowledge/[ePistolarium](#);
  - b. Epistolary metadata + prosopographical data: Cultures of Knowledge/[EMLO](#).
4. Evaluate results.

## WG3: Texts and Topics

### ABSTRACT

WG 3 will complement the strategies for assembling and exploring data on correspondence and correspondents pursued in WGs 1–2 with study of digital means of engaging with the texts of letters themselves. This requires agreeing standards for the presentation of texts both as images (of manuscripts or printed books) and in digital form (I). It also requires developing tools to aid transcription, annotation, and collaborative editing, and for transforming digital editions into print (II). For exploring vast quantities of highly fragmentary textual material, text mining and topic modelling must be deployed (III). Since tools useful to scholars can only be developed by studying scholarly working methods, the Action's Training Schools will be adapted to the purpose of such study, alongside the induction of a new generation into emerging tools and techniques (IV). WG 3 is also the place in which broader discussion of how the tools developed in all the WGs can be assembled to create an integrated and user-friendly 'Virtual Research Environment' (IV).

WG 3 led by **Charles van den Heuvel**, Professor of Digital Method in Historical Disciplines, University of Amsterdam; Head of the Research of the Group 'History of Science and Scholarship' at the Huygens Institute for the History of the Netherlands in The Hague; Senior Researcher in the Virtual Knowledge Studio for the Humanities and Social Sciences at the Royal Netherlands Academy for Arts and Sciences (KNAW). He also plays a central role in the pioneering DH project, '[Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic](#)', responsible for the most advanced experiment with the application of IT to topic modelling of large and multilingual corpora of correspondence: the [ePistolarium](#).

### AGENDA

#### I. Fundamentals: Images of documents, text coding

1. The point of entry of many letters into a digital resources is as *images of manuscripts and printed books*. Shared infrastructure requires standard means not only of publishing such images but also of handling them. Early modern letters normally consisted of one or more sheets of paper, folded to provide their own addressed envelope, sometimes tied with ribbon and often sealed with wax. Due to the cost of paper and postage, writing was often densely packed into every available area, including vertically in margins and even at right angles across previously written text. In allowing users to manipulate images of folder and double-sided paper easily, Shared Canvas promises to provide a valuable tool worthy of close study.
2. Another basic precondition of shared texts and tools is uniformity in the *presentation and encoding of digital text*.

#### II. Transcription, Annotation, and Editing

1. For generating digital from printed text, *Optical Character Recognition* is of steadily increasing utility. WG 3 will seek to pool knowledge on current capabilities and future refinements of OCR. Also explored will be the possibility of using crowd-sourced corrections to inform intelligent computing solutions to enhancing that capability for major early modern typography. A status report on plans to adapt such techniques to manuscript transcription is also needed.
2. For manuscript text, a variety of *transcription tools* have been developed. WG 3 will collect, study, compare, and report on the best options for future use and further development. Means for dealing with multiple *transcription standards* also need to be devised.
3. Given the extremely allusive nature of communication between frequent correspondents, annotation is needed to make learned letters readily intelligible. *Annotations tools*, suitable to individual and collaborative research, need to be assembled therefore and assessed as the basis for further development.
4. Annotated transcriptions evolve into *critical editions*. An assessment of the existing states and future prospects for editorial platforms is also needed.
5. Although searchable and changeable digital text has many advantages, the fixity, stability, and permanence of print offers countervailing attractions. A study should also be made of packages designed to ease the *transition from digital to print media*, including both pedagogical applications and print on demand.

### III. Textual analytics: data mining and topic modelling

1. One of the most valuable features of letters to the scholar is their tendency to contain information on a huge diversity of topics, much of it unavailable elsewhere. The corresponding difficulty is that letters can lurch unpredictably from one topic to another, making the identification of relevant material difficult. One means of rendering large quantities of letters more readily navigable is via *text mining*, which applies a variety of analytical strategies — including lexical analysis, pattern recognition, association analysis, and visualization — to extract information from unstructured texts.
2. Even more closely adapted to scholarly needs is *topic modelling*. In theory, this technology should allow researchers to identify material on specific topics and general areas, even in the largest and most fragmentary collections of correspondence, based on the frequency with which certain terminology pertaining to the topic sought. In practice, corpora in several languages, each with its varying early modern orthography, pose significant challenges. In addressing these difficulties, WG 3 will build on the work already undertaken in the [ePistolarium](#).

### IV. Training Schools

1. Reassembling the republic of letters requires broad-based collaboration. Broad-based collaboration requires the training of a network of scholars in the use of emerging tools,

techniques, and methods. Preparing for future developments requires that this training be directed above all to the emerging generation of scholars in early career. One of the central objectives of WG 3 is therefore to coordinate the *COST-funded training schools* devoted to this purpose.

2. Developing tools which scholars need and will actually use, on the other hand, requires intensive communication between users and system developers. With this necessity in mind, Training Schools in this Action will also be devoted to *assessing scholars' interaction with relevant digital tools*. The first Training School (in 2015) will monitor and assess the needs and experiences of scholars learning to use recently created tools. The second Training School (in 2016/17) will also attempt to assess the future needs of scholarly users with regard to new tools being developed within and outside the community represented in the Action.

## **V. Toward a Virtual Research Environment**

1. Ultimately, all the tools envisaged and developed throughout all the WGs for dealing with the temporal, spatial, prosopographical, social, textual, topical, and physical aspects of correspondence need to be brought together to create a *Virtual Research Environment*, that is, an integrated online interface designed to help a distributed community of researchers collaborate in assembling and exploring this vast and fragmentary literary heritage, in publishing the fruits of their work in a variety of formats, and in projecting the results into the classroom and into the broader public domain. Work on planning such a VRE — one of the culminating objective of the Action as a whole — will also be coordinated by WG 3.

## WG4: Documents and Collections

### ABSTRACT

The principal **aims** of WG 4 are two-fold. One is to contribute to the refinement of a shared data model which includes common definitions of the physical features of the letter, its basic genres, and its modes of dissemination and preservation. The second is to plan and pilot a census of correspondence collections across Europe, and the infrastructure designed to support it.

The **leader** of WG 4, Dr Elizabethanne Boran, is Librarian of the [Edward Worth Library](#), Dublin, and the editor of the [correspondence of James Ussher](#), Archbishop of Armagh (1581–1656).

### AGENDA

**I. Refining the data model.** Amongst the main features of letters and collections requiring careful definition are the following:

1. physical characteristics of letters (watermarks, folding, seals, ribbons, etc.);
2. stages in letter composition (draft, fair copy, copy sent; autograph and scribal version);
3. genres of learned letters (epistles dedicatory, letters to the reader, formal testimonials, official correspondence, newsletters, etc.);
4. reciprocal relationships between letters and other documents and media (oral discussions, minutes, diaries, commonplace books, printed books, news media, scholarly journals, enclosures);
5. modes of dissemination (forwarding, circulation, and reading by recipients; recipients' copies and extracts);
6. modes of collection and preservation, including:
  - a. personal collections, auto-archived by correspondents themselves, with or without related material assembled by others (e.g. [Leibniz-Archiv](#), Hannover);
  - b. named collections, assembled by individual or corporate collectors (e.g. [Waller Manuscript Collection](#), Uppsala);
  - c. institutional collections of official, semi-official and related correspondence (e.g. [Royal Society of London](#));
  - d. letters preserved with non-epistolary material, whether manuscript or print (often overlooked in cataloguing and therefore difficult to track down: e.g. Josten, ed., [Elias Ashmole](#));
  - e. composite collections of manuscript correspondence, composed of several of the above categories (e.g. [Bodleian Library](#), Oxford);
  - f. collections of early modern correspondence printed during the period;
  - g. subsequent scholarly editions of correspondence, in printed or digital form;
7. dispersion and destruction (missing letters; deliberate destruction and theft; accidental or collateral loss or destruction; means of tracking provenance of collections and of individual letters).

**II. Census of correspondence collections.** The necessary precondition for the orderly collection of data on individual learned letters scattered across Europe is a census of the main collections of early modern learned correspondence.

Sourcing collections-level descriptions is a complex task, given the existence of relevant materials in manuscript, print, and digital form. A variety of methods must be developed for harvesting existing data and generating new descriptions, amongst which are the following:

1. for existing data on early modern *printed collections of correspondence*:
  - a. assemble existing bibliographies of early modern printed letter collections (e.g. Estermann);
  - b. extract collections of printed letters from national historical bibliographies (e.g. STC, Wing, VD 16, VD 17, USTC, etc.);
  
2. for existing finding aids (print and digital) for *manuscript correspondence*:
  - a. assemble a bibliography of printed catalogues of early modern manuscript correspondence;
  - b. consolidate understanding of scope and method of existing union catalogues of relevant mss material (e.g. CEN, Kalliope, e-manuscripta);
  - c. assemble links to electronic catalogues and findings aids for major relevant manuscript collections (e.g. Leiden UB);
  
3. for the scholarly crowd-sourcing of *new catalogue-level descriptions*, infrastructure and workflows must be devised and piloted, along the following lines:
  - a. design webform to facilitate the efficient, collaborative inputting of collection-level descriptions of printed and manuscript correspondence into a common database;
  - b. experiment with use of this infrastructure for scholarly crowd-sourcing;
  - c. advertise the opportunity to contribute to this database via associations for archives, libraries, special collections, rare books, and relevant academic fields, potentially in each country across the COST Action;
  - d. establish monitoring procedures for quality control and prioritization of data inputting and systems development.



## WG5: Data Exchange and Strategic Planning

### ABSTRACT

WG 5 aims to negotiate principles and standards of data exchange as well as plans for digitizing and making freely accessible learned correspondence from the early modern period. Building on the survey of collections of printed and manuscript letters, inventories, and finding aids undertaken in WG 4, WG 5 will develop a master plan for a joint digitization program. In addition, the WG will address issues of Open Access and legal restrictions, content syndication (e.g. to EMLO, Europeana, etc.), the development of worldwide unique persistent identifiers for letter, the connecting letters through semantic web techniques, and the long-term archiving of digitized letters.

WG 5 is led by Dr [Thomas Stäcker](#), the Deputy Director of the [Herzog August Bibliothek](#), Wolfenbüttel, responsible for the library's programme of digitization and the [Wolfenbüttel Digital Library](#).

### AGENDA

#### I. Generating digital metadata

The census of collections of early modern learned correspondence (pursued in WG 4) should be complemented by investigation into the most efficient and reliable methods of *generating large quantities of digital catalogue records*.

1. For generating metadata on *uncatalogued collections*, WG 5 needs to explore the rapidly developing area of *crowd-sourcing* — scholarly, semi-scholarly and otherwise. [Zooniverse](#) is piloting flexible software for this purpose, and is looking for humanistic projects on which to pilot it (Oxford to lead).
2. For *existing catalogue data in traditional card files* and related formats, experience of the most cost-effective and reliable means of *scanning and keying* should be pooled (library community to lead).

#### II. Unifying metadata standards

1. This Action devolves the negotiation of individual components of the data model to the specialists assembled in WGs 1–4. WG 5 will chair the Action subcommittee within which these individual components will be integrated into a comprehensive standard data model.
2. Since individual letters can exist in multiple manifestations (drafts, copies, extracts, abstracts, etc. in manuscript, print, and digital form), a *unique identifier scheme for individual letters* is needed. This project should build on the experience of similar schemes for people (e.g. [VIAF](#)) and printed books (e.g. [VD 16](#)).

#### III. Digitizing learned correspondence

The census of correspondence collections (produced by WG 4), could also provide one basis for drawing up a *master plan to digitize collections of learned correspondence*.

1. Collections of *printed correspondence* now out of copyright can be digitized and web-mounted in the manner pioneered by [CERA](#). WG 5's initial role is to recommend standards and best practice.
2. Collections of *manuscript correspondence* can be digitized and web-mounted.
3. The challenge of *funding* such an enterprise could be assisted by drawing together information on various relevant funding schemes at the European, national, regional, civic, and institutional levels.

#### **IV. Sharing digital metadata: legal agreements and scholarly conventions**

Generating large quantities of standardized digital metadata should be coupled with a campaign to persuade those in possession of such data to share it.

1. Part of this challenge is *legal*: we need to develop a range of *Open Access policies* for metadata, images, documents, and full digital editions (drawing in this case on the experience of [CC](#), [Open Data](#), [GNU](#), etc.).
2. Another part of this challenge relates to *scholarly conventions*: we need to develop concise and accurate *citation standard* for digital materials. This might involve employing a persistent scheme (such as [DOI](#), [URN](#) or [Handle](#)), as well as [XPath](#) or [Xpointer](#) techniques) to cite more granular portions of text. A particular challenge is that, unlike printed materials, digital resources change over time, for which versioning pointers might be developed.

#### **V. Recruiting contributions**

With such arrangements in place, WG 5 will coordinate a campaign to encourage contributions of relevant metadata, images, texts, and editions from a range of potential contributors.

1. *Repositories* contributing metadata render their collections more visible and discoverable.
2. *Publishers* of copyrighted editions of correspondence may regard digital catalogue records as advertisements for their products.
3. *Collaborative research projects* gain access to the digital tools and larger pools of data available on shared infrastructure.
4. *Individual researchers* render contributed data accessible and future-proof at minimal trouble and cost.

#### **VI. Preservation and sustainability**

Issues of long-term preservation and sustainability require careful consideration.

1. One *technical challenge* is to develop means to allow central repositories of digitized data to be regularly up-graded and up-dated without contaminating the data or disrupting the functionality of the digital tools developed to process them.

2. Another is to develop an ontology cloud which allows data models, standards, and authorities to be incrementally refined over time as well. Here the experience of Aalto's [Semantic Computing Research Group](#) will be particularly valuable, linking this strategic strand with work being undertaken in WG 2.
3. The *financial challenge* is to develop funding sources and mechanisms which allow the preservation and up-grading of both data and platforms to be sustained indefinitely.

## WG6: Visualization and Communication

### ABSTRACT

The principal objective of WG 6 is to explore the application of visualization and interaction design to the challenges involved in standardizing, browsing, analysing, and understanding the unprecedented quantity of epistolary data and metadata potentially brought together on the digital infrastructure envisaged by the Action. After studying the results of previous work and the strategies for co-designing a next-generation research environment (I), the core task of WG 6 will be to conceptualize and specify a new generation of visualization tools applicable to all stages of the process of working with epistolary materials (II). The group's secondary task is to consider the enhancement of this core infrastructure with means of communicating both the scholarly and the technical interest of the Action to a variety of different audiences, within (III) and especially beyond (IV) the academy. These four main tasks are outlined in more detail in the agenda below.

WG 6 is led by [Paolo Ciuccarelli](#), an Associate Professor at the [Department of Design](#) of the [Politecnico di Milano](#) (Italy) and Scientific Director of its [DensityDesign Research Lab](#).

### AGENDA

#### I. Existing tools and approaches: an overview of previous work

1. collecting and analysing available visualization tools identifying possible collaborations with other institutions and projects;
2. developing co-design strategies to be applied in future collaborations;
3. collecting possible case studies within the COST Action to be used in first prototype/experiments;
4. outlining a general strategy for the design of a future digital platform/environment (together with the other group and especially with WG 3 and WG4).

#### II. Conceptualizing and specifying new tools and approaches

1. helping scholars express and articulate their own experiences with the use of visualization tools and, more in general, with the relational dimension of the data;
2. defining how interaction design and data visualization can contribute to the process of uploading, refining, and enriching the data, with a particular reference to RDF;
3. defining how interaction design and data visualization can participate in the process of exploring the data with a particular interest on the interpretative nature of humanistic modes of inquiry.

#### III. Communication within the scholarly community

1. coordinating the communication and the online presence of the COST Action, notably in the design of a highly interactive Action website;

2. designing, specifying, and costing the creation of a scholarly networking platform;
3. designing, specifying, and costing the creation of an on-board, born-digital scholarly journal, integrated into the union catalogue, and devoted to new kinds of scholarly publication building directly on both large archives of data and metadata and on interactive and animated visualizations of them.

#### **IV. Communication with broader publics**

1. discussing the multiplicity of publics (archives, libraries, museums, universities, schools, publishers, and other cultural heritage institutions) which might be served by the digital infrastructure at the heart of the Action and the enhanced forms of visualization, communication, and interaction facilitated by it;
2. outlining diverse strategies and tactics needed to build proactive and effective reciprocal relationships with those publics, including potentially new forms of digital publication and dissemination.