

- **Overview of goals and objectives of your DiD project**
- The ChartEx Project is developing new ways of exploring the full text content of digital historical records. The project is using medieval charters (title deeds), in machine-readable text, from the 12th to the 16th centuries and focuses on recovering and reconstructing the complex relationships between people, sites and events described within them. There are three core modules:
- Natural Language Processing (NLP): for the analysis and annotation of individual records in English and Latin.
- Data Mining for the discovery of new relationships between entities of actors, sites and events across multiple records.
- An interactive virtual workbench that will allow historians to explore the information extracted, particularly relationships between entities, to reason about those entities and to add further information and comments.
- **Challenges and lessons learned from international collaboration across disciplines and domains**
- ChartEx is a highly ambitious and innovative project. Time and effort funded has probably not been sufficient for the computational science. This is particularly true for the NLP and the workbench modules. Considerable effort has been expended beyond that funded by DiD. Nevertheless the project has kept to the brief of the original bid and identified clear goals for further development in new initiatives.
- More contact is needed between teams in different countries to develop a clear mutual understanding of the difficult interdisciplinary issues and technical details. The lack of a required overall project management structure (shared across national teams) can be challenging. We addressed this by writing a project plan with a clear overall project structure and agreement on objectives, responsibilities and outputs. This was required by JISC for UK partners but would be good practice for others.
- Within the overall plan a modular structure is important so that partners were not entirely dependent on others to carry out all tasks.
- We recommended use of a shared online workspace (for sharing documents, progress reports etc) and face-to-face meetings supported by VOIP meetings to establish common vocabularies, aims, objectives and sharing of outputs. It is still hard to prevent important business being conducted by private email.
- Not to underestimate the difficulties and time in arriving at a shared understanding, even of quite simple terms, concepts and processes, across disciplines. Project glossary of shared terms is useful.
- **Digital humanities, social sciences and computational based research methods in the context of big data projects**
- Digital humanities collections, especially from the distant past, may be relatively small in volume but may be more complex in their composition.
- Digital humanities data in the domain of History may be more 'fuzzy' and less complete. There are additional linguistic challenges relating to historic language use. This requires novel and inventive approaches giving the opportunity for 'real research' in disciplines such as NLP.
- Big data success therefore depends on a critical and close attention to small details in the data (such as nuances and ambiguities in expression, eg names of persons). Big data success supports small data agendas.
- Illustrating that innovative systems can and should be developed from the needs and tasks that historians wish to do, not just from available technologies that historians then need to adapt themselves to – i.e. moving from technology-push to historian-pull.
- The system is trained on relatively modest samples (in big data terms), but is eventually applied to tens of thousands of documents. Having this large collection of documents automatically structured and linked opens up true possibilities for big data science.
- **Indicators of success**
- We have developed an effective markup schema, which has allowed both the training of NLP software and provided an initial data set to test Data Mining approaches. Development of a successful markup schema for English and Latin texts suggests that we have a schema that can be adapted to even more languages.

- We currently (August 2013) have preliminary results on person and site-matching from the DM and a working prototype of the workbench which the historians are beginning to validate and evaluate with the main focus of evaluation planned for the Autumn/Fall 2013.
- Strong interest in further development from Archives and HEI sectors. (TNA (UK), Borthwick Institute, Columbia Digital Humanities). Use of software in other projects.
- Invitations to participate in further research projects and proposals (eg DID3 x 2; AHRC Care for the Future: York/History, ...), AHRC Big Data: *Linking Individuals* (TNA, KCL, Leiden, ...others); ...
- Positive reaction to conference presentations. These are partially caught by our twitter feed (@ChartExProject). Notable examples include: @wgthomas3 19 Jul @mwidner 27 Nov ; Brett Bobley @brettbobley 19 Jul. Write ups on community blogs include: <http://magistraetmater.blog.co.uk/2013/07/10/chartex-data-technologies-and-charters-16226717/>; <http://dh4ph.com/2013/07/23/chart-ex-the-medieval-charter-excavator/>; <http://dmlbs.wordpress.com/2012/12/07/digital-collaborations/>; <http://www.medievalhistories.com/tag/chartex/>
- **Measuring impact**
- Impact on efficiency of medieval scientists. We were supporting the historians to do their work quicker and on a larger scale. Early validation promising.
- Incorporation into archival staff training programmes (TNA; DEEDS, Toronto)
- Incorporation into graduate training programmes (TNA, DEEDS, Toronto, Columbia, York (both History-planned and HCI, Computer Science-current)
- Number of historians (outside those involved in ChartEx) who express interest in being involved in evaluations of the systems
- **Knowledge dissemination mechanism and tools**
- Our dissemination strategy is aimed at engaging domain specific audiences (particularly medievalists) as well as digital humanists.
- Website: www.chartex.org Twitter: @ChartExProject
- To date eleven conference presentations have been delivered at six international conferences 2012-13, with four more currently planned for 2013-14.
- Past conferences: all presentations and powerpoints available here: <http://www.chartex.org/publications.html>
- Future conferences and events: <http://www.chartex.org/events.html>
- Press: all participating University press offices, The National Archives press office. Press releases and links here: <http://www.chartex.org/publicity.html>
- Nine publications are planned by various members of the team in a range of domain specific and digital humanities' journals.
- **Importance of working with libraries, archives and data repositories**
- Such organisations supporting ChartEx include TNA (UK), Borthwick Institute for Archives (UK), Columbia Digital Humanities, DEEDS and University Library Toronto, Dictionary of Medieval Latin from British Sources (University of Oxford).
- **Capacity building and training (students and highly qualified personnel)**
- Most of the RAs employed have reflected on what they have learnt from participating in the project. Reflections include learning about interdisciplinary and international collaboration (including challenges), the gains to be won from painstaking, detailed work ('monks' work' in Dutch), greater insight into the linguistic and technical aspects of data content, developing a more interrogative approach to data, learning very new skills and vocabularies and being excited and surprised (as well as challenged and daunted) at the outcome of the collaboration. A commitment to moving DH beyond digitization.
- At Brighton, Columbia and York graduate students, (masters and PhD level), from beyond the project have engaged profitably with it as part of their research training.

Report contributors: Sarah Rees Jones, Arno Knobbe, Helen Petrie, Christopher Power, Adam Kostó, Michael Gervers, Lynne Cahill, Roger Evans, Ruen-Chuan Ma, Jon Crump, Stefania Perring, Robin Sutherland-Harris.