The place of models and modelling in Digital Humanities: some reflections from a Research Software Engineering perspective

Arianna Ciula, Geoffrey Noël, Paul Caton, Ginestra Ferraro, Tiffany Ong, James Smithies, Miguel Vieira

Sommario
Il presente contributo illustra una posizione sul ruolo dei modelli e della modellizzazione nel contesto del Research Software Engineering (RSE) nelle Digital Humanities (DH). La nostra ipotesi è che all'interno del contesto nel quale opera il King’s Digital Lab, influenzato da fattori storici e di gestione manageriale, la produzione di modelli è di fatto il contributo essenziale dell'RSE all'epistemologia di DH. Tuttavia, riteniamo che quando considerate da una prospettiva di RSE olistica, che include - pur non essendo ad essa limitata – la modellizzazione empirica e predittiva, queste attività di modellizzazione non siano state studiate in modo abbastanza sistemativo da supportare a pieno questa posizione. Il nostro contributo può soltanto indicare qualche idea e abbozzare un approccio che meriterebbero ricerche più approfondite.

Abstract
This paper offers a position on the role of models and modelling in the context of Research Software Engineering (RSE) in Digital Humanities (DH). Our hypothesis is that within the context that King’s Digital Lab operates, shaped by historical and management factors, the production of models is arguably the core contribution of RSE to the epistemology of DH. However, we argue that, when analysed from a holistic RSE perspective, encompassing but not limited to empirical and predictive modelling, these modelling activities have not been studied systemically enough to support such a claim in straightforward ways. Our contribution can only gesture at some ideas and sketch a framework that would need further research.

1 Context
It is important to consider the institutional milieu which informs this paper. The authors currently cover different roles at King’s Digital Lab (KDL), a

---

1 A preliminary version of this paper was presented at the symposium Computational Text Analysis and Historical Change, held at Humlab, Umeå University (Sweden), 4-6 September 2019 as well as other venues and then expanded for the ADHO 2020 Digital Humanities conference (cf. Ciula et. al 2020).
Research Software Engineering (RSE) team hosted by the Faculty of Arts & Humanities at King’s College London. In operation since 2015 and officially launched in November 2016, KDL is a Digital Humanities (DH) lab with significant research software engineering capacity. The team provides software development and infrastructure to departments in the Faculty of Arts & Humanities while also collaborating with a range of external partners in the higher education and cultural heritage sectors (King’s Digital Lab 2018).

The team includes research analysts and software engineers, designers, and systems managers. The engineering team is supported by a project as well as a lab manager, director, and deputy director.

According to Hans Radder “experimentation involves the material realization of an experimental process (the object of study, the apparatus, and their interaction)” (Radder 2003: 4). In the case of KDL, the experimental process is realised in its Software Development Lifecycle (SDLC). As the diagram in figure 1 illustrates, the SDLC guides the interactions within KDL’s experimental process. For the experimental process to take place at all, experts are needed. Expertise entails roles, professional identity, careers but also tacit knowledge and stability (cf. Smithies 2019). The ‘apparatus’ in this case are the technical systems considered adequate for KDL’s experimental process; these range from the lab’s infrastructure architecture to technology stack, to the range of tools and apps used for development (as well as for project management and monitoring purposes).

---

2 As outlined in King’s Digital Lab (2018), KDL is a research infrastructure in the sense that:

We design and implement systems, infrastructure, tools, and processes needed to produce a heterogeneous range of high quality digital scholarly outputs (from historical databases and digital editions to data visualisations and infographics, from georeferenced maps to natural language processing workflows and immersive experiences).
The diagram describes the lifecycle of a project idea evolving from an initial contact with partners to the final release and maintenance phases. Key milestones in KDL projects are easily mappable to Agile dynamic systems development method (DSDM) – characterised by strong foundations and governance – and its project phases (cf. Agile Business Consortium 2014), namely pre-project, foundations and feasibility assessment. They run from initial contact to the first kick off meeting, followed by an evolutionary development phase with planned deployments and release process, followed in turn by the post-project phase (cf. Ciula / Smithies 2023). KDL adopted a tailored version – under continuous improvement - of Agile DSDM to manage priorities within and across multiple projects and to acknowledge the fluid nature of the design and development processes as the solution and the research context evolve during projects lifecycles (cf. King’s Digital Lab 2018).

As argued elsewhere (cf. Smithies / Ciula 2020; Ciula / Smithies 2023), the core intellectual challenges – entangled with technical ones – KDL faces lie in aligning the four dimensions of the lab’s multifaceted socio-technical ecosystem: team, data, models, and systems. The range of interactions amongst the team, systems, processes, and products of modelling at a given moment in time are guided by pragmatic alignment between the social reality of the laboratory environment, technological innovations, standards, digital methods and relevant conceptual frameworks.
2 Definitions: Digital Humanities, modelling and models

KDL operates within a unique context. It claims its origins in the pioneering work of colleagues at King’s College London working in applied computing in the Humanities, beginning in the 1970s (cf. Short et al. 2012). However, the crossings between RSE and DH communities at King’s and internationally are only recently being highlighted and explored (cf. Gold 2009; Smithies 2019; DHTech Group 2019). We argue that the study and building of models is one of the dimensions via which these crossings emerge more vividly, with substantial epistemological implications and innovative ramifications for the field of DH as a whole. We intend models here in a relatively wide sense, as artefacts of different kinds including but not limited to computational models. We focus mainly on data and interface models as opposed to computational and methodological process modelling which are by nature less amenable to being captured in a snapshot and pinpointed. A systematic study of the former can however also reveal insights on the latter.

DH is humanities scholarship transformed and extended by the digital and vice versa. This mutual transformation and extension concerns tools as well as epistemologies (cf. Ciula 2017; Ciula / Marras 2019: 35) and is the object of critical research in DH. DH praxis foregrounds a modelling activity that is present in most research but often unexamined. The core practice of research in DH, as practiced in KDL, is modelling (e.g. cf. McCarty 2005: 20–72; Buzzetti 2002; Beynon / McCarty 2006; Flanders / Jannidis 2018), which implies the translation of complex systems of knowledge or conceptual frameworks into computationally processable models or operational frameworks. The centrality of modelling to KDL is contingent on historical factors - mainly the historical development of DH at King’s, and KDL’s intellectual history linked to the humanities computing tradition - as well as on its first director wanting to position KDL at the intersection of engineering and theory, with modelling being the obvious vanishing point between the two (Smithies 2017).

Findings from a recent project (cf. Ciula et al. 2018) which reflected on definitions of modelling and how they apply to the DH context could be summarised in the following points:
1. Modelling is a process of signification and in particular a creative process of thinking and reasoning, where meaning is made and negotiated through the creation and manipulation of external representations.

2. Adopting Nancy Nersessian’s (2008) continuum hypothesis whereby modelling in research is a refinement of modelling in everyday life, modelling as a research strategy is a process by which researchers make and manipulate external representations - what Peter Godfrey-Smith (2009) calls “imaginary concreta” - to make sense of objects and phenomena.

Such a definition of modelling spans the whole range from ‘deformative’ to empirical modelling and is inclusive of formal or predictive modelling (cf. e.g. Joslyn and Turchin 1993). Models are external representations created and manipulated for specific purposes; they are the residues of modelling processes which make them potential objects of study in philosophy of science but also in other disciplines. From this wide perspective, models are indeed documents which can be studied across the history of the humanities tradition also prior to the introduction of computational modelling.

KDL assumes, therefore, in the tradition of applied computing in the humanities and histories of the humanities such as Rens Bod’s, that modelling

---

3 Cf. what James Smithies argues with respect to DH modelling as an example of postfoundational epistemology and methodology (Smithies 2017: 174-9), and to what he labels “software intensive research” in the humanities:

[...] data analysis in the humanities functions along a continuum of practice, from rigorous studies that ask to be critiqued in formal mathematical terms, to studies in corpus linguistics that use data analysis to build empirically verifiable arguments, to heuristic explorations that aim to surface unexpected information from a large dataset or ‘deform’ a text in unexpected ways to prompt new readings of it. At this end of the continuum, algorithmic analysis is used as a hermeneutic rather than empirical method. (Smithies 2017: 168).

4 With this respect, Rens Bod claims the following:

Modelling is ubiquitous in the humanities: while scholars do many things, the search for patterns and principles, and the links between them, is found in all humanistic disciplines and periods. Modelling in antiquity consisted mainly of explaining and constraining patterns by means of principles. In the early modern period, modelling also included the prediction and refutation of patterns by means of these principles. Since the late nineteenth century, the focus shifted to interpreting and criticizing patterns by means of principles [...]. The exploration of different modelling strategies and practices in the (history of the) humanities has just begun and may lead to a new field coined History and Philosophy of the Humanities (HPH), analogous to History and Philosophy of Science (HPS). (Bod 2018: 78).

For the historicity of models in DH see Knox (2012), Ciula (2017a) and Ciula / Marras (2016).
is *more than* an epistemological or methodological choice for RSE in the humanities, but an unavoidable requirement for the design and development of digital research outputs. It inhabits the space where engineering method and critical theory meet and in this sense is an ontological prerequisite or methodological ground of DH software engineering (cf. Smithies 2017: 3). If it is de-centred from the software development process (in favour of, say, historical method or literary theory) the purpose and nature of the DH research process is fundamentally altered, and the resulting outputs potentially undermined.

In the context of experimental settings applied in multiple disciplines and domains, David Gooding (2003) claims that computational approaches are analogues to other processes of abstraction, measurement and contextual interpretation, whereby reduction of complexity is followed by expansion in the guise of a double funnel-shaped process (cf. Gooding 2003: Figure 13.4; McCarty 2018: Figure 1). Complex phenomena are reduced to enable quantification and manipulation to then be subjected to a process of expansion, for example via the visualisation of findings to make them interpretable. The process of signification that unfolds in modelling activities (cf. Ciula et al. 2018) implies translation, negotiation and transformation of meaning. These transformations occur both in modelling processes engaged with abstraction of complex phenomena into rule-based procedures - what Gooding calls reduction (2003: 280) - and in modelling directed at the re-integration - or expansion (Gooding 2003: 278) - of the results of that reduction into interpretative frameworks such as explanatory diagrams and data visualisations.

3 Modelling in an RSE context

Operationalisation makes models formalised into, for example, snippets of code or software components, but the process of abstraction of target historical and cultural objects or complex phenomena into rule-based procedures is contingent and cannot be reduced to strict formalisation only. Other approaches and languages besides translation into code play a role in the process and, possibly more fundamentally, the process itself is situated (e.g. dependent on personal styles, team shared knowledge, institutional setting) and historical (e.g. dependent on legacy systems, institutional history). With respect to non-verbal languages in an RSE context, the process of data modelling, for example, is often informed if
not driven by communication and collaborative reasoning around more or less standardised graphical representations and notations (e.g. sketches on the whiteboard, diagrams expressed in conventional graphical languages such as UML) in phases of reverse engineering as well as exploratory design methods. Equally, the re-integration or expansion of modelling efforts into interpretative frameworks usually rely on verbal and visual language to document code, or to explain the results of an experiment:

[…] in empirical, technical or highly implementation-focused settings, formal (e.g. computationally rule-bound) and informal (e.g. sketches, narrative, story-based) models coexist and interact to give sense to our modelling efforts. (Ciula / Marras 2019: 39)

3.1 Software Development Lifecycle and lab methods

Going a bit deeper into the practices of modelling in an RSE context, the objects of study of KDL experimental processes are datasets of different kind:

While all data is “taken and constructed” (Drucker 2011), depending on the project idea KDL is engaged with, data sometimes already exists at pre-project phase (often in need of massaging or cleaning), or, in other cases, data is collected as part of a project activity. Either way, the team makes sense of project data and imports it into KDL systems (when not already present in some legacy format) by constructing and interpreting models. (Ciula / Smithies 2023).

The products of this experimentation process are models, which from an RSE perspective fulfil

[…] the pragmatic function of bridging the tension between idiosyncratic objects of study, research questions, methods and contingent datasets on one hand, and scalable and sustainable solutions on the other […] while standards and technologies change, evolve, and decay, the effort to achieve adaptable models continues. (Ciula / Smithies 2023).

As mentioned above, we intend models here in a relatively wide sense, as artefacts of different kinds including but not limited to computational models. They can be produced during several phases of the SDLC. Thus, as outlined in Ciula and Smithies (2023), modelling can encompass, for example:

- negotiations around the meaning of the project units of analysis documented in diagrams and definitions which shape an agreed project
language (during a design process that usually starts in the pre-project phase and evolves throughout the lifecycle - cf. the neologism repeteme in Salciute Civiliene et al. 2019);

- paper or whiteboard sketches used to draft the solution architecture for a project (design method in feasibility or foundations phases - cf. figure 2);

- wireframes and static mockups of user journeys (design method in evolutionary development phase - cf. figure 2);

- data models implementing the logical structure of a database (build method in evolutionary development phase - cf. figure 3);

- statistical models implemented with ad hoc algorithms and code or relying on tested formulas and existing libraries (build method in evolutionary development phase - e.g. logDice measure for lexical association devised by Pavel Rychly (2008) and adopted by Pak (2018)).

Figure 2: Whiteboard drawing, conceptual design map, and static mockup by Ginestra Ferraro in collaboration with Miguel Vieira and Arianna Ciula, for the Ego-media project (under final review).
Figure 3: Data model by Geoffrey Noël in collaboration with Eleonora Litta and Elena Pierazzo for the KDL legacy project *Early English Laws* <https://earlyenglish-laws.ac.uk>.

Figure 4 below serves to illustrate the key high level operational methods that shape KDL SDLC phases, namely design, build, maintain and monitor methods (cf. Smithies / Ciula 2020). Modelling activities occur as part of design, build and maintain methods.

Figure 4: Coloured interrupted lined rectangles cluster overarching key methods in KDL SDLC (by Ong, based on Smithies / Ciula 2020: fig 3).
We intend and use design methods in a wide sense ranging from techniques of requirements elicitation in pre-project analysis to data modelling and wireframing in evolutionary development. Therefore, by design methods we include what are often considered two separate activities in design and development process studies, namely design and analysis. This is because from a modelling perspective the two are connected and it is by modelling that analysis of a domain of knowledge and its description moves more or less smoothly into the design of the specifications for a new system. Building methods are also subsumed within design activities in this diagram and represent the points in time when design moves from abstract to concrete forms and converges into development products. From a Software Engineering perspective, data models have a cascading effect on storage solutions, interface design and data integration (cf. Pasin and Bradley 2015). Therefore, while models affect intimately and substantially the published collaborative digital outputs usually submitted to national research assessment framework (cf. Ciula 2019), from an RSE perspective, these are often the tip of the iceberg of the experimental process.

3.2 RSE roles

Partially documented within SDLC templates (cf. King’s Digital Lab 2018) and other material (e.g. from narrative and files in code repositories to sketches, tasks, notes, discussions and governance documents in project management tools), modelling occurs at every phase of KDL SDLC as well as across its operational methods. One of the core functions of modelling is to support the translations of cycles of analysis and design (both part of KDL design methods). Therefore, not surprisingly, modelling also occurs across RSE roles whereby:

Creativity and openness to diverse research domains, as well as the ability to identify patterns of similarity across heterogeneous projects, also play an important role, and are integral to RSE expertise and processes. (Ciula / Smithies 2023).

---

Note that this is in line with computer sciences and information systems design approaches which tend to see analysis and design as complementary activities that contribute gradually to elaborate models as a project progress (cf. for example Bennet et al. 2005).
Each modelling cycle produces one or more models which can contribute to bridge building phases and increments (analysts ← models → developers ← models → designers). This process is far from linear and unidirectional. Its epistemological value can be limited to one role (for example an analyst sketches a model of a domain of knowledge for her own understanding and to inform requirements elicitation at a later stage), to more than one role within the engineering team or indeed to the overall research team including partners outside KDL. It facilitates communication, shared understanding and ultimately the building of a final product charged with meaning sedimented in more or less ephemeral intermediary products (widely defined here as models). More often than not, models are also shared outside the research team of a project with other users and researchers (e.g. in focus groups and workshops as part of user research and testing or dissemination activities).

Figure 5: Diagram by Tiffany Ong: models are produced and shared/re-worked within the team and with external contributors.

Beyond instantiations in specific time-limited projects, in the RSE context where they were designed and developed, models often represented innovative solutions which have had a longstanding effect - think, for example, of the factoid model explained in Bradley (n.d.) and adopted in a series of prosopographical projects at King’s College London, or the model
for the components of handwriting (to which Paul Caton contributed) underpinning the Archetype framework (Stokes 2011) for digital palaeography initiated by the Department of Digital Humanities at King’s, and now being used in further research and tools development internationally. The sustainability of models depends on an interlocking chain of factors, of which, in the case of KDL, context, standards adoption and code integrity are paramount. They are codependent on core elements of the lab’s socio-technical environment: namely expertise, systems and data.

4 Sketching a framework for RSE modelling journeys in DH

Others have attempted to conceptualise the multidimensional space of modelling activities in software engineering. In particular, Gonzalo Génova et al. (2009) propose to see modelling activities trajectories in a three-dimensional space along the axes of purpose (whether the model is a description or the specification of something e.g. of a new system to be build), reality (whether a model is representing a domain of knowledge or a system), and abstraction (whether a model is more abstract or concrete). Based on this useful conceptualisation, the examples in figure 2 above represent a series of models to define the specifications of a system moving from the abstract to the more concrete. However, as transformations in project development spearheaded by Agile approaches have demonstrated, RSE modelling trajectories are all but linear. The organic evolution of these trajectories in KDL SDLC makes the intersubjective role of modelling emerge. What follows are two examples taken from the Community of the Realm of Scotland project (s. https://cotr.ac.uk/), where the definition of what a dynamic edition entails evolved from conceptualisations by project partners and discussions during the foundations phase. These resulted, amongst other things, in the rephrased - translated in a different language of expression and transformed as meaning was made in the process - definition summarised in the diagram (cf. Broun et al. 2017-21: network graph) produced by the project analyst (Caton). This abstract model describing (and firming down) the domain of knowledge was complemented by a glossary (cf. Broun et al. 2017-21: dynamic edition concepts). The two are to be read interdependently.
During a project lifecycle, modelling journeys evolve non-linearly with moments of disruptions and breakthroughs in tandem with the development of the project intellectual and practical outputs. In the case of the COTR project, one of these moments was the translation or transformation of the previous phases of modelling into an Entity Relation data model by the project software engineer (Noël). This diagram contributed to define further the conceptual space of the project and in particular to move the project design and development towards a more concrete footing which via other transformations evolved into the Django (web framework) models available on the KDL code repository at https://github.com/kingsdigital-lab/cotr/blob/master/cotr/ctrs_texts/models.py.
The last stop of the modelling journey at time $C_{n+1}$ in figure 7 is a logical database diagram which software engineer Vieira - who joined the project after several increments and releases - generated automatically to understand the system and see it represented with a convention he was familiar with. This example illustrates both the entanglement of analysis and design as well as the non-linear bidirectional nature of RSE involvement: the implementation imperative (design) requires a level of details that triggers questioning (analysis) and thus ripples and reshapes the project partners' conceptual domain.

Despite their conciseness, these examples aim to illustrate at least four components of our argument in this paper:

- The idiosyncratic and distinctive way DH evolved at King’s and in particular at KDL where modelling is positioned as epistemologically and methodologically central to its strategy and operations;
- The co-production of models across SDLC phases (from foundations to evolutionary development in the case) and RSE roles (models produced by system manager and designer in the project are not included here to simplify the journeys, but those occurred too);
- The intersubjective role of models in facilitating knowledge sharing and meaning making around a project research questions within the
RSE team and beyond (in particular in negotiating language and defining the project conceptual space);

- Models are subject to non-linear transformation and hence bear creative potential to bring in multiple perspectives, new knowledge and understanding as well as to foster innovative technological solutions. As exemplified, this transformation takes place across languages of expression (verbal descriptions, graphical representations following more or less standardised conventions, code).6

We can imagine several ways by which our modelling practices across SDLC phases and roles could be improved. Here we focus on how modelling practices could be improved at the operational level, raising their research and intellectual value. Firstly, we believe an ethnographic study of KDL’s practices could provide a space for observation and reflection but also inform operations at a practical level, for example by developing an adequate format for modelling notebooks evolving during the SDLC and across roles.7 Secondly, while the literature on modelling spanning disciplines and traditions is wide and varied, KDL has only recently passed its start-up phase to be able to reflect on and distil a language around modelling adequate to its own context and activities. Working mainly within a Digital Humanities context offers the opportunity to give prominence to the critique of and narrative around models. These reflections are an integral part of the research projects within which they emerge but are not often explicitly documented and tend to be reconstructed post-factum. Supported by a systematic ethnographic study, documentation of modelling activities could follow several strands: for example, from the adoption of machine-readable and standard formats for models and enterprise architecture metamodels, and frameworks such as The Open Group and NAFv4 Architecture Framework (NATO Architecture Capability Team 2020) to better integrate design workflow and guide the definition of increments in the development process.

6 We did not dwell on the mapping to standards in the field (e.g. crosswalking), but it is another obvious transformation process models are subject to in the RSE context.

7 A 2-year Marie Curie fellowship was awarded to Urszula Pawlicka-Deger to conduct training and ethnographic research at KDL from October 2020 (see <https://dhinfra.org/>).
Ultimately, our perspective is informed by an inclusive understanding of who modellers are and what they do to contribute to the research process within collaborative projects. Agency is therefore to be shared and empowered across RSE roles and including project partners, for example by adopting tools that facilitate the co-editing of models relying on shared manipulable visual languages (something made more urgent by the remote working mode during the COVID-19 pandemic). Last but not least, while better internal documentation and integrated process are key to sustain and nurture personal and collective tacit knowledge within the team, exposing modelling practices and considering their products at the same level as software would be beneficial for the large Digital Humanities community, including professionals working with digital technology in the arts and humanities, cultural heritage and cultural industry sectors.

5 Alignment with pragmatic and critical modelling

When analysed both from an insider’s RSE perspective as well as from a wider perspective which looks at the extreme offshoots of Gooding’s double funnel, models contribute to define and redefine objects of study which come charged with layers of scholarship and analysis, with previous selections, bias and political as well as ethical responsibilities. As the creators of new memory regimes and intermediaries to the past engaged in modelling efforts which interact and affect the materiality of our objects of study (cf. Ciula 2017a), we bear myriad responsibilities (cf. Ciula 2017b). Discussion around the representativeness and constraints of the digital archive have been raised by librarians, philologists and historians in the DH context (cf. e.g. Dahlström 2010; Hitchcock 2013; Prescott 2014). More recently, Katherine Bode (2018, 2020) presented some lucid analysis around modellers’ responsibilities in digital literary studies by exposing the gaps that propagate from literary works we know of, to material preserved in the analogue archive, to selections of works that make it into digital archives, to further reductions in the creation of a corpus of analysis and in the application of statistical modelling techniques. The latter dictate additional powerful yet limiting constraints if not contextualised critically within an interlocking chain of bias (see figure 8).
Within KDL’s experimental context, models are nothing but the documents, the artefacts or the residue of more or less ephemeral modelling processes; our contribution aimed at reflecting on how models produced within cycles of design and development come into being and what effect they have in the resources they contribute to instantiate and hence in interpretative processes of expansion as well as reduction. The framework we propose is aligned with a “critical modelling” approach (cf. Bode 2020), but with the addition of material culture and media literacy perspectives, whereby models are seen as artefacts and therefore studied as objects with a multidimensional biography. Our framework pays particular attention to the following dimensions:

- historicity and subjectivity embedded in models (cf. Ciula 2017) (note that this is different from a relativistic stance; cf. Bode 2020);
- iconicity (cf. Ciula / Eide 2017) and representativeness (cf. Prescott 2014) which explains the relation of models to their objects or targets (or domain) in terms of explicit, implicit, manual as well as automatic processes of representation, selection, gaps identification, introduction of bias as well as reasoning potential;
- politics and ethics of models (cf. Bode 2018, 2020);
– performativity of models (cf. Gray 2018) which focuses on what one can do with models making emerge their imaginative, creative, and transformative potential.

In line with these loose directions, the questions which could guide the study and use of “pragmatic modelling” (Ciula / Marras 2019) are therefore people-centred: What is a model for? Made by whom? For whom? In what relation with its target object/s? What does and does enable with respect to reasoning and actions?

Bibliography


The Open Group <https://www.opengroup.org/> [12. 02.2021].