TOWARDS A PROCESSUAL PERSPECTIVE ON ARCHITECTURE

Building an Information Infrastructure for Personalized Medicine

Margunn Aanestad

Johan Ivar Sæbø

Thomas Grünfeld

University of Oslo, Oslo University Hospital

Abstract:

In this paper we discuss the emergence of architecture, understood as the structured relationship between components in suites of software systems. We aim to formulate a processual perspective on architecture, implying that we emphasize the process of emergence over time as opposed to the usual time-invariant, static views that dominate architecture discourses, i.e. architecture as process as opposed to architecture as a model. We also intend to foreground the actual practices of "architecting", emphasizing architecture as an accomplishment rather than as a given. Finally, we specifically focus on the handling of temporal issues in this process. Architectures are crucially related with the longterm perspective and are built upon predictions of the future, but in reality they are also shaped through compromises between long-term and short-term demands. As our empirical case we describe the emergence of a new information infrastructure for personalized medicine, where we describe the emerging configuration of elements and component and how they were put into a structured relationship. Theoretically, we draw on the notion of "shearing layers"(Brand, 1994) used to denote subsystems with different rates of change. We show how considerations about expected rate of change was a significant factor shaping the emergent architecture in the case study.

Key words: architecture, information infrastructures, personalized medicine, genetics

1. Introduction

This paper studies a future (but currently emerging) information infrastructure for personalized medicine. This is not a simple application, but a multipurpose, multilevel and multi-stakeholder information infrastructure. The planning and realization of such infrastructures is challenging as it does not only have to address a single purpose, a single level (as in "level of the stack") or a single or a few stakeholders. The field of personalized medicine is also one that is changing rapidly, and where there are a lot of future unknowns. We wish to describe the evolution of this with an emphasis on how decision-making during the design, development, implementation and scaling phases reflects the need to accommodate multiple stakeholders, purposes and to address simultaneously several "layers" in the infrastructure. In doing this, the perspective of the developers oscillated between the visionary future (the long-term concerns) and the existing present situation (the short-term concerns).

The reason for our interest in these processes is that we believe there will be implications for scalability when design is driven by local and immediate concerns, as well as implications for generalisability when design is driven by particularity of (local) needs. We think it is significant to examine whether (and how) choices of diversifying and/or generifying the solution remain open over time, or not. This view implies that the system, and the underlying constellation of components and their relationships, is dynamic and fluid. Applying the concept of "architecture" as it is used in the information systems literature, we find that this concept is too static, and gives the impression of a steady state vision. An environment where the knowledge base and understanding of the addressed topic is rapidly developing further underscores the irrelevance of an "initial blueprint". The objective of this paper is thus to further our understanding of IS architecture as more dynamic, opposing much of the existing literature. This is also an answer to a call for a redefinition of the research agenda on IS architecture (Bygstad and Pedersen 2013).

Before we present the research approach and the empirical findings, we briefly present architecture as used in the information systems literature, and the domain of personalized medicine to provide the necessary background.

2. ARCHITECTURE IN THE INFORMATION SYSTEMS LITERATURE

The metaphor of architecture is widely used in practical work with information systems, but the theme has received less attention as a research topic. The metaphor is generally used in various ways and about many 'levels' or 'objects', e.g. software architecture, information architecture, enterprise architecture, business architecture, security architecture, technology architecture etc. Architecture is "a plastic concept" (Scheil, 2008), it is "usefully ambiguous and is often used at a high level of abstraction" (Bidan et al., 2012), and may therefore "serve as a shared boundary object…between various stakeholder groups engaged in systems development (Smolander et al. 2008). In a literature survey, Vassilakopoulou and Grisot (2013) identified three discourses around the notion of architecture. The first is a technically oriented stream that primarily addresses software architecture. The topic of interest in this stream is to develop and evaluate various architectural approaches, such as architectural frameworks. A core challenge that these frameworks deal with is the need to deal with different 'views' or perspectives of the multiple stakeholders in the project, each of which a temporary and limited description (see e.g. Zachman's (1987) classic framework). The frameworks and process models that have been developed are helpful for mediating and communicating between the participants in the process, thus, their deployment may support the organizing necessary for the construction of a complex artefact. A second stream of research adopts a more explicitly socio-technical approach, and discusses isses related to e.g. business and enterprise architecture. Architecture is here seen as a way for managers to impose control over a messy reality, and the concerns of the researchers and practitioners within this discourse is more related to how one can manage the processes, e.g. how one can monitor adherence to 'the blueprint' throughout a project. A third research stream is labelled «architecture as strategy» by Vassilakopoulou and Grisot (2013). The researchers in this stream targets architectures and approaches to defining and implementing architectures as their objects of study. This stream sees architecture as relevant because it is a central approach to sensemaking, action, and governance in the face of complexity. This literature typically acknowledges the emergent and improvised nature of architectures, and does not share the previous stream's "architecture as control" ideas. The necessity, benefits, challenges and experiences of governing a system or an information infrastructure through generic principles such as decoupling, decomposing, modularity, layering etc. are in focus in this research, within which our study falls.

Architecture is a metaphor that influences our thinking, it "shapes the categories, discourse and language used" (Scheil, 2008) and it "helps us to reason about a system" (Bass et al., 6). The metaphor of architecture is taken from the domain of constructing physical buildings, and the traditional usage of the notion of architecture emphasizes the structural qualities of a set of elements: which elements are included, the relations between them, and the properties of the relations and the elements. This usage of the metaphor foregrounds the static and spatial quality of architecture. However, as is evident in the evolution of multiple architectural approaches in the industry, developing blueprints and models is not enough. Many frameworks also encompass a process support component, based on the realization that defining, implementing and maintaining architectures require also the organization of the process. These process frameworks (such as TOGAF, OIO, DoDAF) therefore also specify the sequence and steps, together with representations and documentation along the way. In this paper, we want to build on such a processual view on architecture. This goes along with a view on architecture as practice, or what we could call "architecting" (in a parallel move as when researchers study "organizing" instead of "organizations"). It is interesting to note that the notion of architecture (as a noun) derives from the role of the architect (and not from a plan or model for a building). The word architect has evolved from Greek *arkhitekton* which is composed of *arkhi*- (chief) and *tekton* (builder), i.e., the chief builder, master builder, director of works. To study the actual work practices of the 'architects' in practice, will help us to see what "architecting" really is.

The processual nature and the practical accomplishment is a core point of departure for the paper, and we believe it is a useful angle for theory and practice. The definition of a target architecture is not in itself ensuring that it will be implemented, and in practice, usually a transition strategy from the 'here and now' situation towards the goal situation needs to be defined. This may be widely experienced in practice, however, it is not foregrounded in the discourse. There is a lack of understanding of the nature of how architectures actually evolve, and of how they are governable. We aim to contribute to a better understanding of how architectures evolve, and wish to emphasize the role of organizing temporality. There are several reasons why we believe that this is important. Brian Foote and Joseph Yoder (2000) has described settings in which there is not a high-level, preplanned architecture, but a "Big Ball of Mud" architecture; a "haphazardly structured, sprawling, sloppy, duct-tape and bailing wire, spaghetti code jungle. We've all seen them. These systems show unmistakable signs of unregulated growth, and repeated, expedient repair. Information is shared promiscuously among distant elements of the system, often to the point where nearly all the important information becomes global or duplicated. The overall structure of the system may never have been well defined. If it was, it may have eroded beyond recognition". Foote and Yoder discuss the reasons for the emergence of these kinds of system landscapes, and point to a number of interesting observations that we wish to address. Two of the most crucial challenges of realizing well-ordered architectures are related to the inherent tension between long-term and short-term concerns, as well as to the need to adapt and learn along the way. The first challenge emerges from the fact that architecture intends to address the long-term concerns of a system. In practice, short-term issues such as deadline approaching may challenge the architectural concerns which may have to yield. Architectural concerns may be seen as optional, like a costly luxury that takes up resources better used for other things. Secondly, "architecture is a hypothesis about the future" (ibid., p. 6) and Foote and Yoder emphasize the learning required for building good architectures. They claim that sub-optimal and immature architectures may be good as temporary measures for not-so-well understood domains, since they allow data and functionality to migrate to their natural place. Premature architectures may be worse than anything, since they fix the architecture before enough is known about the domain. To accommodate the need for learning, flexibility is required, and architectures must be evolvable. This however, may cause conflicts with the need for stability and consistency. One way to resolve this is through classic principles of modularity and decomposition according to guidelines such as “low coupling and high cohesion”. Another is to apply a layered architecture, where the layers are relatively decoupled. This allows the layers to evolve with different rates of change, something which is captured by Brand (1994) by the notion of “shearing layers” in a building. Brand proposed the existence of six layers: Site (the geographical location), Structure (the load bearing elements), Skin (the exterior surface), Services (the circulatory and nervous systems of a building, such as its heating plant, wiring, and plumbing), Space Plan (walls, flooring, and ceilings), and Stuff (includes lamps, chairs, appliances, and paintings). The point is that these layers change at different rates. Similarly, software artifacts experiences different demands to changes, and some elements of a system may have to adapt at different rate than others. Thus, in well-working systems, architecture emerges where the rate of change is more or less consistent within a “layer”, and where interaction happens within a layer or between adjacent layers.

3. GENETICS AND PERSONALIZED MEDICINE

“Personalized medicine” denotes the usage of knowledge about a patient’s genetic profile to indicate the appropriate treatment, such as selecting the most efficient drug. Currently, there is a certain degree of usage of genetic information in ordinary medicine, mainly for diagnostic purposes (to determine the cause of an illness or defect), and for prognostic purposes (e.g. to examine for the presence of genes known to cause disease in order to assess risk). However, very seldom are genetic tests used as a basis for making decisions relating to the treatment. There are initiatives to introduce elements of this in cancer treatment (e.g. the UK’s Stratified Medicine initiative as well as Norwegian initatives) where the patients’ tumors’ DNA are examined. The therapy approach (surgery, chemotherapy or radiation etc.) is decided based on the genetic profile of the tumor or genetic background. Another usage domain for personalized medicine is pharmacogenetics, where the selection of drugs and determination of appropriate dosage can be based on genetic profile.

The evolution of the basic technologies for sequencing of the DNA has a large impact on the unfolding of these scenarios. The introduction of High-Throughput Sequencing (HTS) of DNA allows for drastically faster and less expensive sequencing. This technology is used to analyze the full genome and complements older sequencing technologies that analyse only parts of the genome. The previous, so-called targeted tests, analyze specific genes, for instance BRCA1 and BRCA2 which are the genes known to cause breast and ovarial cancer. High-Throughput Sequencing (HTS) is an established technology in a research context, but only to a limited extent in a clinical context (i.e. in hospitals, health service) where the older sequencing technologies (such as Sanger sequencing) dominate. With the development of cheaper and better HTS technologies, this approach may come to dominate genetic analysis also in a clinical context. For instance, currently in Norway, a blood sample from every newborn are subject to testing for some 40 specific diseases, and it could soon be more efficient to replace this test package with one full-genome analysis that would cover all these tests. If this is done, then the question arises whether it would make sense to store this fullgenome sequence, which does not changethroughout the child's life, in case one would like to query for other issues at a later point in time. The same argument for storing the genome data would apply when a full-genome test is conducted later in life for adults.

If the usage of genetic information should be broadly available and integrated into medical practice, a series of challenges needs to be addressed on the ICT-side. A multi-layered information infrastructure for personalized medicine needs to be established. A basic infrastructure is required to run the computation-intensive actual sequencing and variant calling. After a list of variants are generated, specialists need to analyse this list, sort between more or less significant variants and provide a clinical report that offers the required conclusions related to e.g. whether there are variants that are verified or probable causes for disease, or whether the findings are inconclusive etc. This work is today primarily non-automated and many genetic laboratories have long waiting times. There is thus a need to increase capacity of the interpretation in the laboratories. This may happen through rationalization and partial automation of this analysis process. This work involves interactive consultations of international databases, resources and tools, so the infrastructure should offer flexible and disseminated access to these while upholding security. Also, in this interpretation process, there may be a need to consult colleagues, which may be located elsewhere, so a way to share data and information is required. The output from the genetic laboratories is then fed back to the clinical department that requested it. Currently, a minority of clinicians use the available genetic services. For this to increase, the knowledge about relations between variants and diseases needs to evolve, the clinical processes and procedures needs to be updated, as well as the individual clinician's knowledge about genetics.

If the demand for genetic services increase, also the infrastructure required for communicating requests and responses needs to be improved from today's manual practices (Wright et al 2011). A way of digitally communicating requests for tests, as well as communicating the results needs to have appropriate mechanisms for identification, authentication and authorization. As the knowledge about genetics increase, one can envision that some queries are simple enough to be automated. In the field, visions of some variety of 'expert system' are therefore common. However, the domain is both ethically, professionally and legally contentious, genetic data are sensitive and in principle non-anonymizable. Current regulation stipulates for instance that one cannot disclose prognostic information to a patient without genetic counselling, and there are legal-ethical controversies around e.g. how to deal with 'the right to not know', or what to do with incidental findings in a practical context (i.e. what to do with findings that you did not look for). The building of future information infrastructures in this domain is therefore characterized by uncertainty.

4. RESEARCH METHODS AND DATA COLLECTION

We are involved as researchers in an ongoing R&D project where the aim is to establish an infrastructure that could facilitate increased use of personalized medicine in a clinical context. The partners are Oslo University Hospital and University of Oslo, where the authors are also employed. We have participated in meetings and discussions in the project group, and in addition the three authors have been involved in different ways in the practical work of the project: One of the authors is the project manager, coordinating efforts across the various disciplines involved, another has conducted a number of interviews with the project members, and a third author has conducted interviews and observations of work in the hospital (among potential users). Other project members are involved in legal and ethical frameworks, coding of various modules, working on secure hardware and communications.

We thus report from an ongoing and dynamic project where we ourselves are active in shaping the development trajectory, together with a diverse team of both medical, legal, and informatics background. The case description in the following section is thus based on our involvement in this team. In addition, for this paper we have conducted a retrospective analysis of project documents, including the original application, the yearly formal reports to the funder (the Norwegian Research Council), our own notes from project meetings and steering group meetings, in addition to notes from meetings with external stakeholders, presentations to various audiences, and white papers produced within the project.

In light of our active direct involvement in the project, we must stress that our findings and conclusions are our own interpretations of the described events. Codifying interview transcripts and project documents will necessarily involve us interpreting and labelling human practices post-hoc. Nonetheless, we believe that interpretive case studies lend themselves to generalizations and theory-development (Walsham 1995).

5. THE STORY OF ‘ARCHITECTING’ IN THE PROJECT

5.1 Initial framings and demarcations of the project

The project application to the funding agency was written during January and February 2011. The application text argued for the need to establish a national ICT infrastructure for secure and disseminated access to HTS-based genetic information that would be stored for later queries. The project was thus future-oriented and visionary, and the initial framing introduced a number of constraints or demarcations: by focusing on HTS it did not intended to address the older sequencing technologies (such as Sanger sequencing and micro-array). The project would address clinical usage and not research purposes, for which the majority of existing infrastructures, solutions and tools were built. It would focus on germline genetics (i.e. variants in the inherited DNA) and not cancer genetics (which analyzes the mutations within the tumor itself in addition to the DNA of the patient’s body), which was the area where the majority of previous personalized (or ‘stratified’) medicine was initiated. Furthermore, the project aimed to build a solution that offered distributed use, not a solution that would be confined to one hospital setting. It was intended for a future situation where full genome data was collected from a large population, and stored in one (or more) centralized repository(ies). While this was not a legally possible model at the present time, there were expectations that one would see changes in law to accommodate for scenarios with more readily available access to sharing patient information between healthcare institutions and health service providers.

The application was granted in May 2011 and the project was formally and practically initiated in September 2011. A steering group with representatives from the partner institutions was established, as well as a working group with staff from the partner institutions. Not all staff were salaried from the project itself, but were employed on similar projects or as support staff. The account we present here is mostly based on the activities within this core working group. The initial months were characterized by mutual learning and sensemaking, as well as negotiating the ambitions of the project, discussing the practical mode of working, such as the frequency and type of meetings, and ways of organizing the team (setting up mailings lists, document repositories etc.). From our own field notes we see that we were uncertain whether we are going to build a "solution", a "platform" or an "infrastructure". In other words, the concrete nature of the undertaking was vague. However, a "strawman architecture" (figure 1) was presented by the project leader as an initial conceptual sketch for thinking. This sketch depicted three core parts, a basic data infrastructure, an expert system and the various user groups.

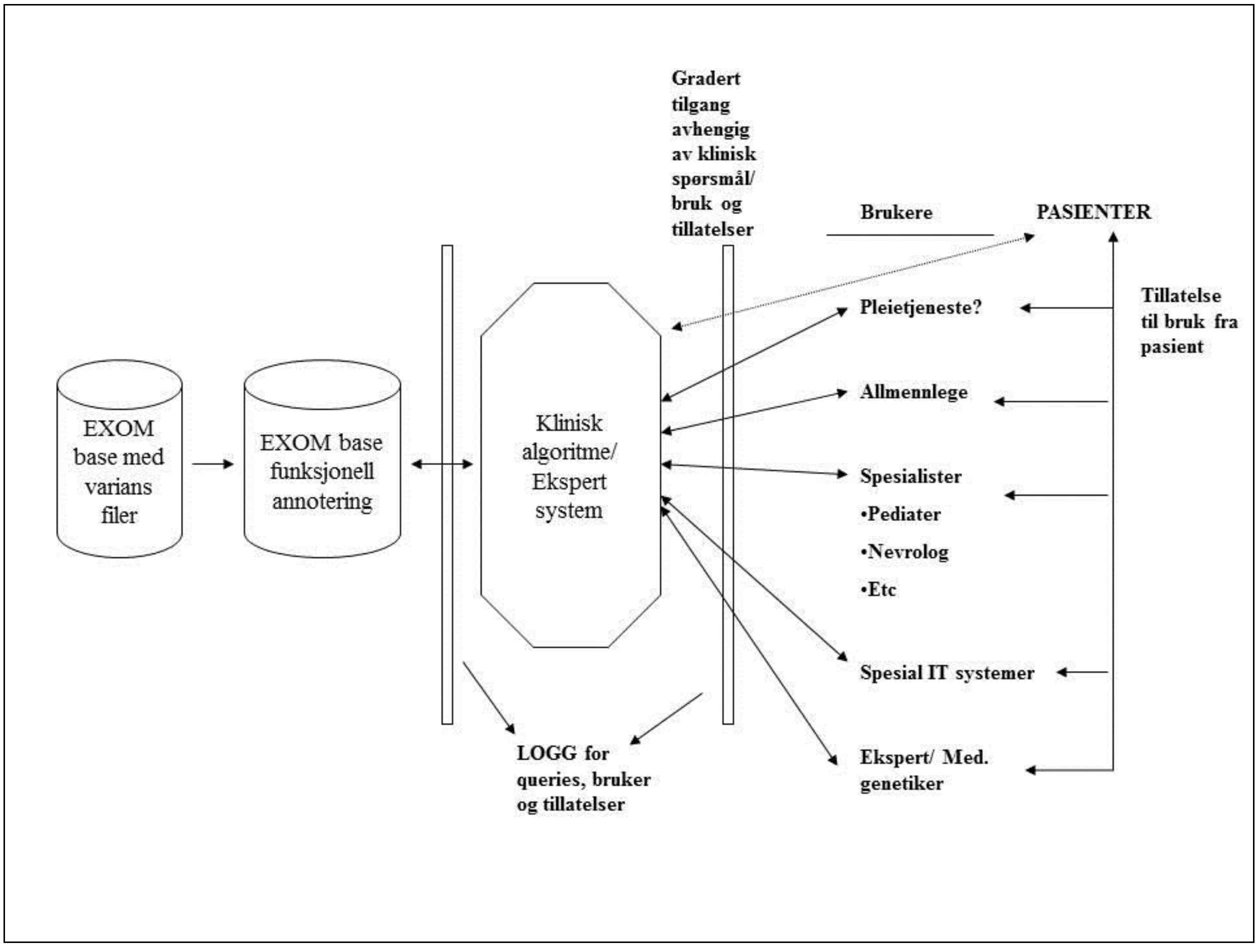

Figure 1 Sketch of "strawman architecture" from September 2011

A number of more elaborate architectural representations were produced as the project continued, and figure 2 indicates one of the most recent ones. This is, however, not yet fully implemented.

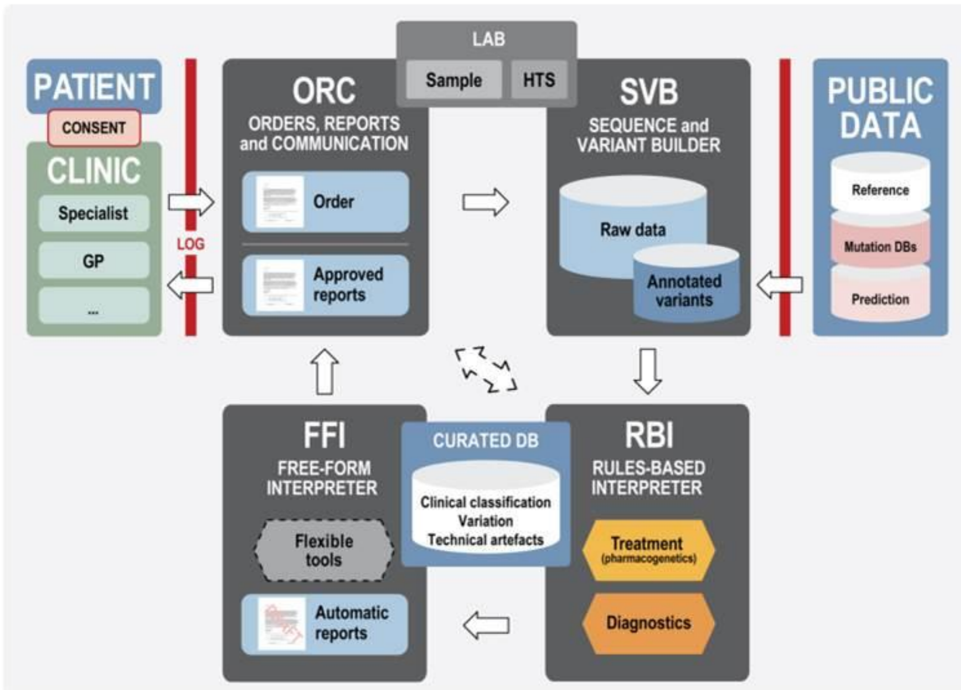

Figure 2 The project's conceptual sketch of the solution per February 2014

5.2 The resulting solution at the current point in time

At the time of writing (September 2014) the solution that is planned to be prototyped is significantly different from the grand visions that were described in the application document and discussed in the early phases. The application which is to be prototyped has emerged around the curated database component that is depicted between the components labelled FFI and RBI in Figure 2. Instead of offering an expert system to doctors in the clinic, the system is a support system for interpretation by the lab staff, implemented in order to increase the throughput and quality of their work. Instead of utilizing only HTS data, the solution can also use Sanger data (the main type of genetic data currently used). Instead of being a shared, inter-organizational solution, the system is (hopefully) going to run within one department in one hospital. Instead of a generic expert system, the implemented rules in the prototype are specific for certain 'gene panels'; picked within relatively simple and well charted domains (monogenic and dominant inherited diseases). Despite the limited actual functionality, the solution is still thought of as a starting point for a more generic and widely used solution. Possibly, such a migration pattern "from the old to the new world" is a prerequisite to make the journey happen at all in a larger organization, where legacy will easily triumph "clean slate approaches".

In the following sections we will describe the various processes that lead to the establishment of the project's ambitions. While it has been an explicit strategy to pilot "small" to demonstrate value, there has also been a downscaling of ambition as to what the pilot and "finished" system should cover. The main storyline will describe a bottom-up emergence of the architecture and describe the way the long-term and short-term concerns were negotiated. In this account we will have to blackbox one major activity in the project, the activities of the University‘s IT center in building the basic infrastructure for secure high-performance computing. This activity proceeded in a relatively decoupled manner from the other activities in the project, however, recently the security demands imposed by the solution have had impact on the progression of the prototyping process. Neither do we describe in detail the activities during the project's first year to build relations to other relevant actors nationally and internationally. The focus will therefore be on the activities within the core group and the implicit and explicit architectural decisions that were made here. We have selected four examples that illustrate core qualities of the process of an emerging architecture.

5.3 Shifting focus from use scenarios to implementation practicalities

The curated database was one of the first components where actual coding started, as it was seen as a core component. It was intended to be the department's knowledge repository of genetic variants, as it should store information on the variants found previously and the categorization (5 categories ranging from causing disease to not causing disease) which had been assigned to them in previous analyses. The database was modelled by a bio-informatician from the hospital department and contained a number of relevant tables that were derived from his knowledge of the usage domain. The initial data model can be said to embody a "patient-centered" organizing principle, since it was derived from the department's work practices. This became evident when the work to implement and populate the table started. When the controller (a business logic layer on top of the database) was programmed, the focus shifted towards questions of how to populate the database, and this introduced a "variant-centric" organizing principle into the design. For instance, the controller accessed only one of the tables that were initially in the data model (the Annotations table), but this was used heavily. A number of additional tables and fields were consequently introduced into the data model and implemented in the database.

5.4 Constructing use scenarios: identifying possible implementation domains

During the first months of the project different possible clinical areas were discussed for implementation of a "demonstrator". A number of criteria were assessed: preferably the knowledge about gene-disease relationships should be mature and well-established in the chosen domain, and this should be familiar to clinicians so that their engagement would be easier. This was important because there had to be interested clinicians that wanted to work with the project on a non-salaried basis. Next one had to assess how well the HTS technology performed for that domain, and a lot of methods verification and validation (of HTS against the older tired and tested methods) had to be conducted. The volume and 'status' of patients affected, as well as the clinical impact of any intervention, determined the practical significance of targeting the domain. The ideal domain would have low-hanging fruits, with low risk and high gain. The first domain to be tested was pharmacogenetics. The presence of two specific variants in a certain gene would predict whether the patient had a normal or a high rate of metabolizing a certain drug. A proof-of-concept mock-up of a decision support system was made for clinicians to help them determine the appropriate dosage of a particular medicine. The potential users' response to the format and presentation of information was studied (Lærum et al. 2014). However, the work in this domain was not pursued, as further activities would hinge on more substantial involvement from the pharmacology department than was possible. More recent research publication also seemed to indicate that the relationships were more complex than initially perceived. On the other hand, the diagnostic procedure piloted in the project is now implemented in the clinic, on a simpler platform. The next domain for a demonstrator was the diagnosis of risk for inherited breast cancer (analyzing presence/absence of mutations in the two genes known to relate to breast cancer). This was a domain where the department of genetics already conducted a substantial amount of analyses (however, based on the older Sanger technology, not HTS) and where existing knowledge was fairly robust. Here there were obvious opportunities to improve the throughput capacity of the department, and the project could align with an ongoing activity of defining Standard Operating Procedures (SOPs). Detailed process maps (flow charts) were produced and a mock-up GUI developed based on the desired work process as described in the SOP. This GUI provided the representation of the application to be built from a front-end, user-centered perspective. Together with a model for the curated database, the representation of the GUI embodied architectural assumptions, notable the GUI represented a process logic that was the point of departure when the controller component was to be developed.

5.5 Learning dynamics and the need to consider "generification"

In the initial version of the controller everything was hard coded, since the new programmer needed to learn to master the relevant tools and frameworks. Thus, the first version was built specifically for breast cancer and had links to the tools and databases used within this domain. However, as time progressed, the code was repeatedly refactored. More and more of the elements that could be expected to change were removed from the source code and put in a configuration file (Such as menus relating to process steps, which external database is accessed, which sequence of steps, etc.). Currently, the gene panel (set of genes examined out of the whole genome) can be changed, as can transcript (functional version of the gene) and the selected external databases that a user will wish to access.

On a larger scale, there had also been discussions about the "generification" of the overall system on a longer time frame. As this was a research project run by hospital employees that simultaneously had a University affiliation, they could utilize the University's infrastructure. However, in the long run, if this should become a routine health service, should the health sector itself establish a High-Performance Computing infrastructure? Or should it integrate with the nation-wide e-infrastructure for biological research? It was not obvious whether or how these will be linked, and the current project strategy is therefore to retract the project resources to building an internal, working component to the future information infrastructure, and let the 'larger' decisions emerge when they do so.

5.6 Having to reduce ambitions due to (external) contingencies:

The HTS technology was the core motivating argument for the application, but it was a novel technology and not yet ready for widespread usage. As the project discussed possible use cases, such as cardiomyopathy and breast cancer, the bioinformaticians in the team attempted to verify whether the HTS technology would yield analysis data of sufficiently good quality for the specific genes and locations that were most relevant. Related to the trade-off between coverage and quality, HTS prioritizes coverage, i.e. it maps 'everything' but not necessarily with a lot of 'depth'. Before a comparison between HTS and established sequencing technologies (such as Sanger sequencing) was established, it would be difficult to convince clinicians to start to use HTS data. It turned out to be more difficult than expected to establish the quality of HTS, and it was said one would have to "wait for better kits" to be provided from the biomedical industry.

The question then arose whether one in the project rather should build solutions that utilized the existing technologies, for which there were data and actual usage in the clinics.

One of the themes in the first project group meeting was on how to integrate the genetic information into the already existing tools of the clinicians, such as the electronic medical records. One of the participants, a medical doctor working in the IT department, was adamant that one could not expect uptake by clinicians if they had to access yet another IT system, so the presentation of results had to show up within the EPR, or at least linked from within the EPR. He proceeded to investigate which solutions for decision support were already in use, what could be other candidates for offering decision support based on genetics, and what standards were relevant. However, the hospital were in a limbo situation with respect to their EPR system, as a few months before the project started, a large and ambitious IT project had been cancelled and throughout the first year of the project period there were no formal decision on EPR policy. In August 2012 a decision was made to exchange the pre-existing systems with one other EPR system, starting in 2013 and finishing in the end of 2014. Therefore, the project was not able to achieve much on this front, as it would be difficult to mobilize resources in the hospital, vendor and service provider organization to build a short-lived integration. Therefore the project activities were concentrated on developing the other parts of the solution.

6. ARCHITECTURE AND PROCESS

The case described above highlights the dynamics between the "here and now" and the "future there", even if we for simplicity have kept several additional unknown variables out of this short account. These dynamics have, following a traditional understanding of the concept, had implications for the architecture of the system. Such a view would give a different perspective of what the architecture is, at any point in time. We will in this section build on this perspective to argue for a processual view on architecture.

Contribution 1: architecture is process, process is architecture:

The project's initial goals depended on a lot of conditions that were not yet in place, such as more mature and widespread sequencing technologies, changes in the legal regulation, more widespread knowledge and demand among clinicians etc. There was both unknown territory that had to be mapped (e.g. in which clinical fields was the current knowledge mature enough that a demonstrator could have practical impact) and territory that needed to be created through the project, such as demand among the clinicians and capacity to deliver more automated genetic analysis and interpretations. This was therefore not a situation where a traditional architectural blueprint was possible to create up front. The system architecture emerged over time and became increasingly concretized and detailed. On the one hand the concretization happened through postponing some of the larger goals. As the exploratory work got underway, technical, legal, and political reasons led to a reduction of the scope of the architecture. For instance, work on generic access control solutions, drawing on external infrastructures for authentication and authorization, was postponed. Similarly, the observation that integration with current EPR systems should not be prioritized within the project's time frame led to a simplification of the information input and output mechanisms designed in the system. Had the situation been different with respect to what could realistically be expected by external partners, the emerging architecture would have become different. On the other hand, the concretization of the architecture happened through a learning process, where the areas that were kept within scope and focused on, were explored further. As they were explored further and increasingly detailed through design and development activities, a lot of learning and adaptation occurred. The merging of basic informatics design principles (as in the original database design) with usage scenarios (embedded in the GUI) revealed the need to modify the intended system. This became clear once the development of the controller software between these two components was initiated. Moreover, the demands of the selected usage area impacted the emerging architecture. The choice of diagnostic work of breast cancer shaped the emerging system into a support system for the genetics department. Had the initial link with the pharmacogenetics domain been pursued, a different system architecture would have emerged, where solutions for end users (clinicians) had taken on a more prominent role than a solution for the interpreters. Therefore we claim that not only does architecture emerge through a process, also we claim that the specific sequence of the process shapes how the architecture becomes.

Contribution 2: importance of temporality

The emerging architecture built upon the existing and utilized the opportunities in the present situation. At the same time there was an attention to the future. Architecture is a "hypothesis about the future" (Foote and Yoder, 2000, p.6) and sometimes the hypotheses are not confirmed and a mismatch between the sketched architecture and the reality occurs. As the architecture in the project emerged, some aims 'scaled off' and were left aside. For instance the aims that involved dependencies on the external actors (national security infrastructures, hospital's EPR projects) were not pursued because one judged that these preconditions would not be in place within the project's time frame. The judgment of the expected rate of change relates to the principle of "shearing layers" proposed by Brand (1994). An expected or perceived difference in rate of change between two domains would indicate that the design should de-couple the two domains. Similar considerations emerged also in other design decisions in the project. The design of the business logic embodied a high degree of flexibility for changes in the information about the technology used, the specific gene panels and transcripts examined, the external databases accessed, etc. Part of this was to accommodate for the rate of change in the field of genetics (both on the hardware and software side, as well as emerging research knowledge on gene-disease relationships), and parts of it was to accommodate for changes that would be introduced by the system scaling. Deployment in a new clinical domain would imply other gene panels, transcripts, databases etc. The expected dynamic character of the usage domain thus impacted the design, and lead to decoupling of the component's relations in the architecture. This can metaphorically be seen as the emergence of shearing layers. The short-term feasibility determined the selection and prioritization of tasks, but still the long-term visions were kept in the project. This is visible in the sketches of the architecture, which kept the representations of a complete system. The long-term perspective was also kept in the design principles employed, not the least in the ongoing refactoring of the code to ensure its capability for reuse and scaling. The selection of a clinical domain for which to develop a pilot was seen as a place from which to start growing the system, not as a representative or core domain to be guiding the whole design. Therefore we may say that the system was built with a perspective of "forward compatibility" as well as "backward compatibility" by necessity, and the decisions would be taken within this extended temporal decision space.

Contribution 3 architecting is organization-building

Our last finding relates to the interplay between the information system and technology, and the organizations that both shape and are shaped by them. This mutuality of organizations and technology is well known, and our case shows how this also unfolds in the architecting processes. The modularity of the emerging information systems implies a need to enrol individuals, groups, and organizations, which hold different skills, experiences, and interests. The conceptual sketch in figure 2 represents not just modules and functionalities, but also current and emerging organizations that hold responsibility for the modules. The case of introducing HTS technologies in clinical medicine underscores this challenge strongly along two dimensions (For a review, see also Hastings, R et al, 2012): 1) Deployment of this complex technology in itself requires cross functional collaborations of experts from different areas of research, including clinical medicine, molecular genetics, HPC IT technology, algorithm development, ethics, social sciences and legal issues, and 2) there is an increasingly blurring border between healthcare and research. Managing a rapidly developing technology, which has obvious major health benefits, but also holds uncertainties along a wide spectre of dimensions (often interdependable), requires not only an "architecting perspective" on the IT structures, but puts similar requirements on organizational development on how to develop, supply and maintain these services. This calls for a dynamic interplay between "IT architecting" and "organizational architecting".

From an additional perspective, it is attempting to reflect the process of a jazz band containing highly trained and professional performers, successfully "jamming" together without ever having played together before. Decision happens real time, but within an overall concept of form and direction, creating the customized and creative end product. A blueprint (here as a music sheet) would be counterproductive or devastating. In a rapidly developing, complex environment, facilitating a productive process may be far more fruitful than initial blue prints and detailed control, - "control is for beginners" (Mills-Scofield 2013).

7. CONCLUDING REMARKS

We have described some insights from an ongoing case study of an emerging architecture for personalized medicine in order to start to formulate a "processual perspective" on architecture. We have emphasized how the architecture was not predefined and implemented, but emerged during a process. We have also shown how the emergence was shaped by multiple mechanisms. The emergent architecture was shaped through a concretization of abstract sketches, through pilots, and through changes to earlier sketched solutions that occurred in a learning process as this was actualized. We argue not only that the architecture was developed in process, but also that the resulting architecture was influenced by the steps and sequence of steps in the process. At the same time we have tried to foreground the role of a temporal awareness in the project and how this shaped the design process. While the need to achieve results on the ground necessitated a "backward compatibility" of the design, there was at the same time an intention to build according to a "forward compatibility" principle as well. This goes beyond a view on architecture as static descriptions, and of implementing architecture as a matter of ensuring compliance with a pre-defined blueprint. We believe that a greater sensitivity to the processual aspects of "architecting" is useful for making better sense of how complex sociotechnical structures in our society actually emerge.

8. REFERENCES

Bass L, Clements P, Kazman R. Software architecture in practice: Addison-Wesley Professional; 2012.

Bidan M, Rowe F, Truex D. An empirical study of IS architectures in French SMEs: integration approaches. European Journal of Information Systems, 2012;21(3):287-302.

Brand S. How buildings learn: What happens after they're built. New York, NY: Penguin Books; 1994

Bygstad B, Pedersen N. (2013) Arkitektur handler om praktisk arbeid i organisasjonen, ikke en tegning. En forskningsagenda om IT-arkitekters utfordringer. NOKOBIT 2013, 1-13

Foote, B. and Yoder, J. (2000): Big Ball of Mud. Chapter 29 in Pattern Languages of Program Design 4, edited by Neil Harrison, Brian Foote, and Hans Rohnert, Addison-Wesley, 2000

Lærum H, Bremer S, Bergan S, Grünfeld T (2014) A taste of individualized medicine: physicians' reactions to automated genetic interpretations, Journal of the American Medical Informatics Association, 14:21, 143-146

Hastings, R et al. (2012): The changing landscape of genetic testing and its impact on clinical and laboratory services and research in Europe. European Journal of Human Genetics, Vol 20, pp.911-16

Mills-Scofield, D: Control is for beginners, Harvard Business review blog network: http://blogs.hbr.org/2013/11/control-is-for-beginners/

Scheil M. (2008) IT Architecturing: Reconceptualizing Current Notions of Architecture in IS Research. ECIS 2008.

Smolander K, Rossi M, Purao S. Software architectures: Blueprint, Literature, Language or Decision? European Journal of Information Systems. 2008;17(6):575-88.

Vassilakopoulou, P. and Grisot, M (2013): Exploring the concept of architecture in Technology and Organization studies. IRIS 2013.

Walsham G (1995) Interpretive case studies in IS research: nature and method. European Journal of Information Systems, 4, 74-81

Wright C, Burton H, Hall A, Moorthie S, Pokorska-Bocci A, Sagoo G, Sanderson S, Skinner R. PHG Foundation (2011) Next steps in the sequence. The implications of whole genome sequencing for health in the UK

Zachman JA. (1987) A framework for information systems architecture. IBM systems journal. 1987;26(3):276-92.