NTNU

Norwegian University of Science and Technology

The genus of information infrastructures:

Architecture, Governance and Praxis

PhD defense by Saptarshi Purkayastha

April 15th, 2015

Agenda

- About me
- Introduce the research
- State of the art
  - Theoretical synthesis
  - AGP model of activities
- Describe the context and setting of research
- Results
- Discussion and evaluation of the model
- Conclusion
- Future research

About me

- For this research, can be considered an ‘involved researcher’, actively involved in the action-research, in development of the artifacts and participating in the research network
- Background in computer science and long-term contributor to open-source projects, two of which are part of my research
- Started the PhD in 2010 in the GHeI – VERDIKT projekt under Eric Monteiro at NTNU and PI, Jørn Braa from UiO, with a break of 1-yr in between and then 4-months on WHO contract
- During the PhD process, moved from ‘an outsider’ in DHIS 2 and OpenMRS to ‘an insider’, with core and leadership role in the OpenMRS community at present.
- Supervisors: Eric Monteiro (NTNU), Kristin Braa (UiO) and Hallvard Trætteberg (NTNU)

About me (2)

- Mostly interpretivist, but a few papers are also on a positivist epistemology.
- Ranging from software engineering, design to implementation science research.
- Need to thank a number of people, naming all of whom is impossible on slides, but people from Oslo, HISP, Trondheim, HISP India, WHO, ministries of health in different countries. Family and friends for their support and kindness.
- Currently, visiting assitant professor at Dept. of Bio Health informatics at Indiana University Purdue University Indianapolis (IUPUI), USA

Introducing the thesis

- Health Information Systems (HIS) are fragmented and results in duplication of work, ineffective use of resources, incomplete and incorrect information
- When data from multiple HIS are brought together, we create an integrated eHealth infrastructure (IeHI) rigged with complexities that are difficult to manage
- These complexities arise from interconnections between socio-technical components of Information Infrastructures (IIs)
- I use the large body of existing literature about IIs and create a taxonomy to classify the activities in IIs. This taxonomy provides clarity to observe II evolution and provides a systematic view to see what activities helped establish the II.
- This thesis includes a collection of 6 papers.

Why is this important (and timely)?

- Previous research has indicated HIS function in their silos and do not talk to each other
- Current thinking at both the global (e.g. HMN & WHO, 2008) and national levels (e.g. in India, Rwanda, Philippines) to address the problem of technical and institutional fragmentation is to implement IeHIs.
- A form of digital infrastructure, IeHIs have been costly, ineffective and non-sustaining endevors in a number of high-income and resource-rich settings.
- What are the chances that these will not be the same in low-resource settings? What can we learn from previous literature about the *success* and *failure* of establishing large, complex IIs?
- Not just from health care, because that is only an instance of a larger set of *infrastructuring* problem.

Research Questions

- Why classify?
  - Better comprehend the complexity of IeHI by dividing the activities into smaller parts.
  - Encourage cladistics to understand evolution of IeHIs
- The thesis attempts to answer the following research questions:
  - RQ1: Given attention to the ongoing efforts of developing infrastructures, how can activities in an IeHI be classified using a taxonomy?
  - RQ2: What are the blind spots created through this taxonomy and how do they affect the infrastructure evolution?

Table: Summarizing papers and contributions (✓-explicit; * - implicit contribution)

	C1	C2	C3	C4	C5
P1	✓ (RQ1)	✓	*		*
P2	✓ (RQ1)	✓	✓ (RQ2)	*	
P3	✓ (RQ1)		* (RQ2)	✓	
P4	✓ (RQ1)	✓ (RQ2)			
P5	✓ (RQ1)	✓			✓ (RQ2)
P6	✓ (RQ1)				✓ (RQ2)

Contributions

- The main contributions are:
  - C1: A taxonomy to classify II *infrastructuring* activities into Architecture, Governance and Praxis.
  - C2: Highlight the organizing categories of activities by analysing cause-effect relationship in cases from eHealth implementations.
  - C3: Articulating Information systems implementation success in terms of meeting local needs.
  - C4: OpenScrum agile methodology to improve knowledge sharing in OSS communities.
  - C5: Defining Big Data through Organizational Capabilities that can be leveraged by the use of Operational BI Tools for analytics in IeHIs.

Theoretical synthesis

- Identified three bodies of literature that focus on different aspects of IIs. Huge bodies of literature with diverse use-cases
  - Information infrastructure ‘core’
  - National/Global information infrastructure
  - Cyberinfrastructure, e-Infrastructure, knowledge infrastructure

- **Information infrastructure core**
  - Infrastructural inversion
  - Standardization and flexibility
  - Installed base and bootstrapping
  - Inertia and slowness of change
  - Few concepts from Actor-network theory, but kept growing, including CAS and other complexity science
  - New concepts like *Control-drift*, *Gestell*, *Reflexive standardization*, *Grafting*…

- STS is the native philosophical home. Groupware, healthcare, oil and gas. Deeper understanding of human-nonhuman than structuration theory.

Theoretical synthesis (2)

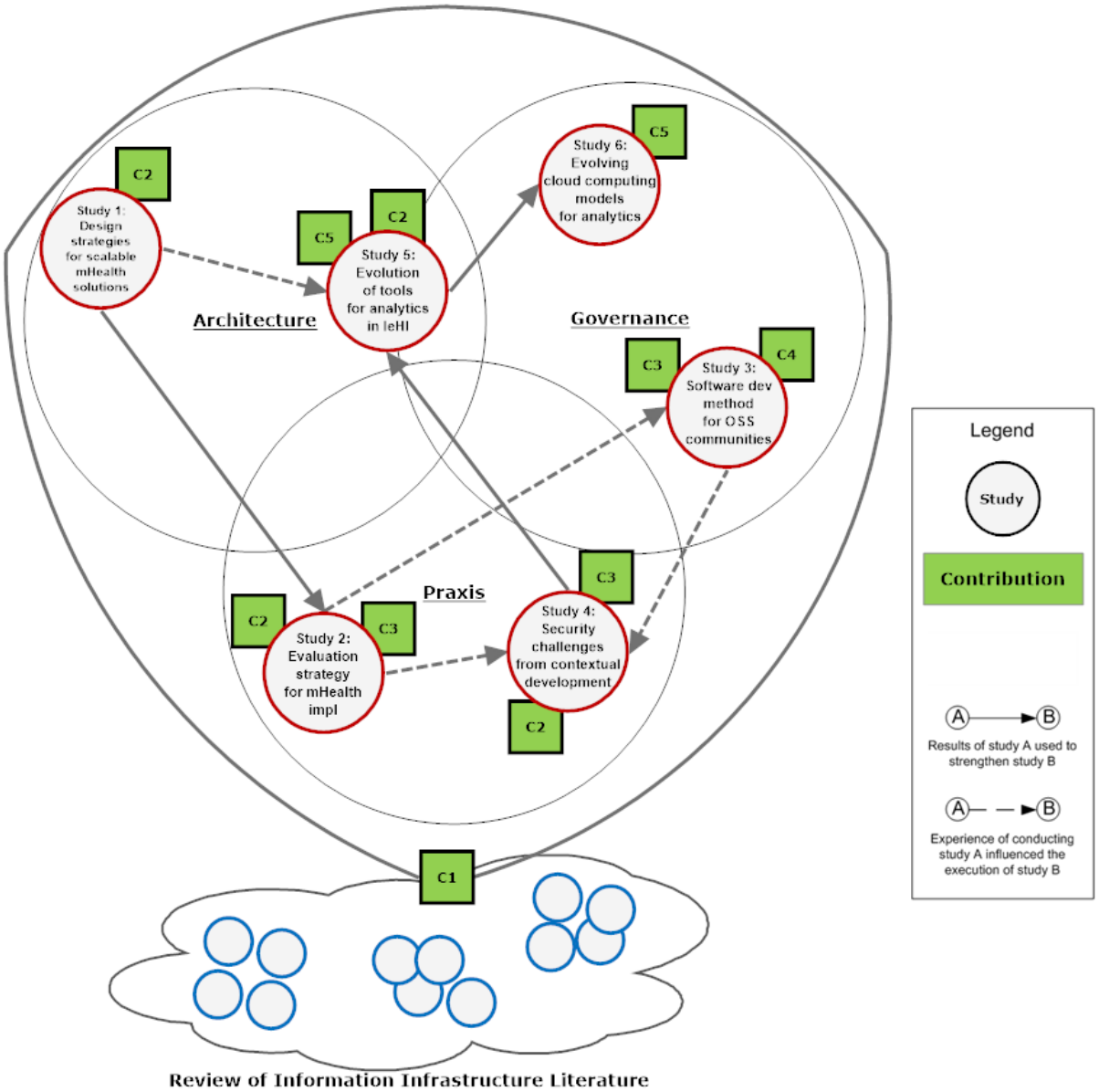
- **National/Global Information Infrastructure (NII/GII):**
  - Policies, laws
  - Social and cultural values
  - Property and Rights
  - Innovation and competitive advantage
  - Equity
  - Networks of action
  - Routines, institutionalization
  - Marginalization
- Politics, policy and power are the main subjects of this body of literature. Broader domains with cases from Internet, defense, GIS, health care, transportation and communication systems.
- More theoretical ‘mix-and-match’ including actor-network, structuration, institutional theories.

Theoretical synthesis (3)

- **Cyberinfrastructure/e-Infrastructures/Knowledge infrastructures:**
  - Alignment
  - Global Vs local
  - Gateways
  - Boundary objects
  - Reverse salient, path dependence
- Subtler theoretical leanings. Less theoretical adoption. More adaptation of complexity science, community of practice and CSCW subfield.
- Main domains are in research infrastructures, use of knowledge systems.

Actor-network theory	**Complex Adaptive Systems**	**Institutional theories**	**Innovation theories**
Translation	Reflexive standardization	Institutionalization	Bootstrapping
Inscription	Path dependence	Counter-networks	Reverse salients
Boundary objects	Variable speed of change	Marginalization	Innofusion
Gateways		Routines	Property
			Competitive advantage
			Equity

Research design



Architecture

- *Infrastructuring* is the work required to establish an infrastructure. This includes the property of *infrastructural inversion*, where the work using the infrastructure is also the work to establish an infrastructure.
- Architecture is often expressed through two ways – (1) using diagrams composed of re-usable or non-re-usable boxes or blocks; (2) using processes and information exchange between the processes and the boxes and blocks.
- Practice of **designing** with a style with regard to **context** – period, place or culture.
- There are also certain specifics in the kind of architecture that can support an II. E.g. open Vs closed architecture, centralized vs decentralized, top-down vs bottom-up, hierarchical control vs user-control.

Governance

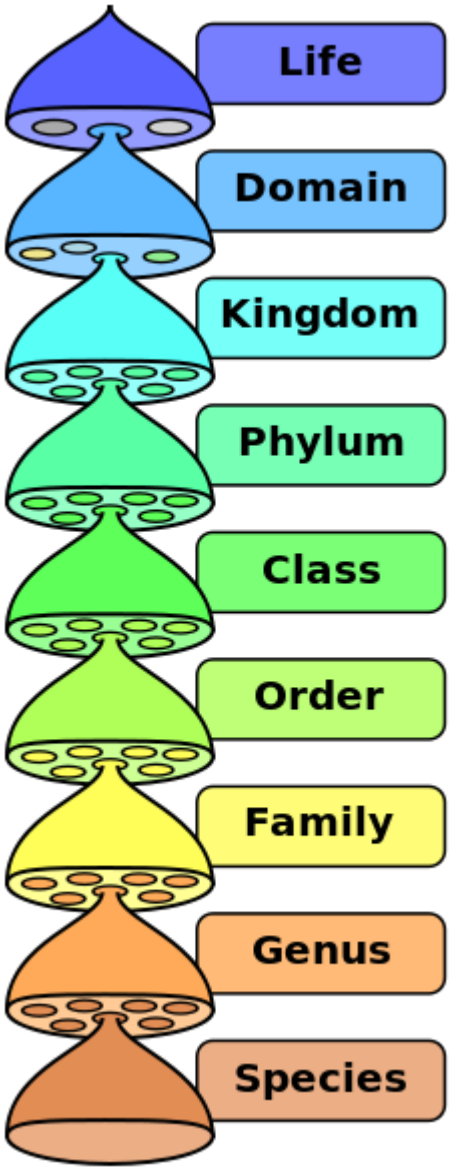
- Governance is largely a mechanism by which decisions are made between participants and the role and activities that they are **supposed to perform**.
- Here *supposition* is important to understand because through governance mechanisms, one only hopes that its participants will act in an envisioned way and formulates laws to punish/reward participants that do-not-perform/perform in the envisioned way.
- Thus, the concept of *Gestell* can very well be understood as the governance principle of providing support to the II by participating in its practices and not “designing” it.
- The GII/NII literature talks about framing policies or governance rules to be able to implement efficiency in the II. Some users have to change their practices more than others.

Praxis

- Enactment, embodiment and actual standards of practice.
- While **Architecture** and **Governance** are visions and attempts at defining and describing how the infrastructure should be used, its use in practice is different.
- *Praxis* in many ways describes the *Bricolage*, which happens in IIs and the challenge of *reflexive standardization* or *workarounds* between global and local work practices.
- Many philosophers describe praxis through an individualistic notion - “human praxis”. Yet there is evidence to *collective praxis* in education (Kemmis, 2010), sociology (Schulz, 1998), psychology (Cahill, 2007). This particularly gels well with human and non-human networks in an II.

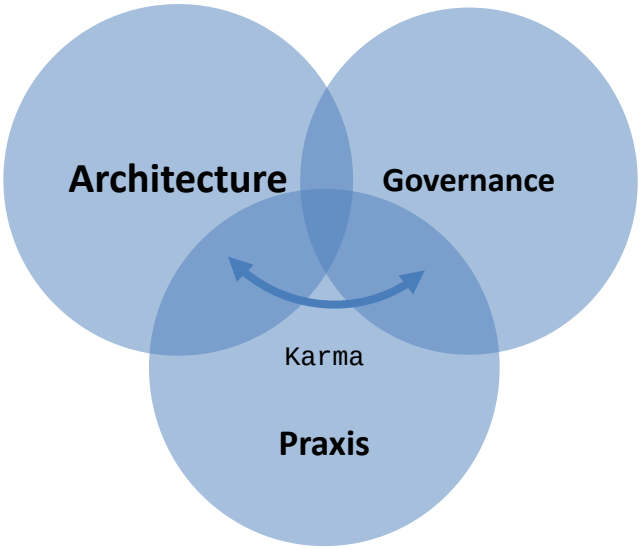
The AGP model

- AGP model is an organizing technique, a taxonomy exercise that is the result of assimilation of bodies of literature.
- It also does not focus on adhering to a specific methodological (quantitative or qualitative), theoretical or epistemological belief like the generative mechanisms focus on being called critical realism.
- Activity from each genera (A-G-P) has effects on the activities in the other two genera.
- Causal powers in the activities possess *sentient causality* because unlike Newton's 3rd law of action-reaction force pairs, *sentient causality* is not deterministic.



Sentient causality or Karma

- Karma comes from the Sanskrit root ‘kri’ which means action, affairs or activity .
- *Karma* is the perception of an individual for happiness or sadness, which is considered outcomes by the doer (actor) of the action and the doer relates these to good actions or bad actions respectively.
- Similarly, actors in an II consider activities to be “stabilizing” or “destabilizing” to the establishing, adoption or scaling of an II.
- Yet these are only time-limited, data-limited world view of the users. Often the words used are “death”, “rebirth” are used to describe infrastructure that fit well with the use of the word *Karma*



Research studies and results


HAPPY NEW YEAR

Study 1: What are some design strategies for scalable mHealth solutions?

- SCDRT mobile application that uses an existing HIS infrastructure of DHIS2.
- Consider the mobile to be "part of a larger whole"

- **Strategy of alignment to existing Infrastructure for mHealth apps**
- The strategy is to first analyse the existing infrastructure and continuously adapt the solution to fit the infrastructure
- Architecture here is a strategic choice of when to use new technology and when to build upon existing infrastructure.

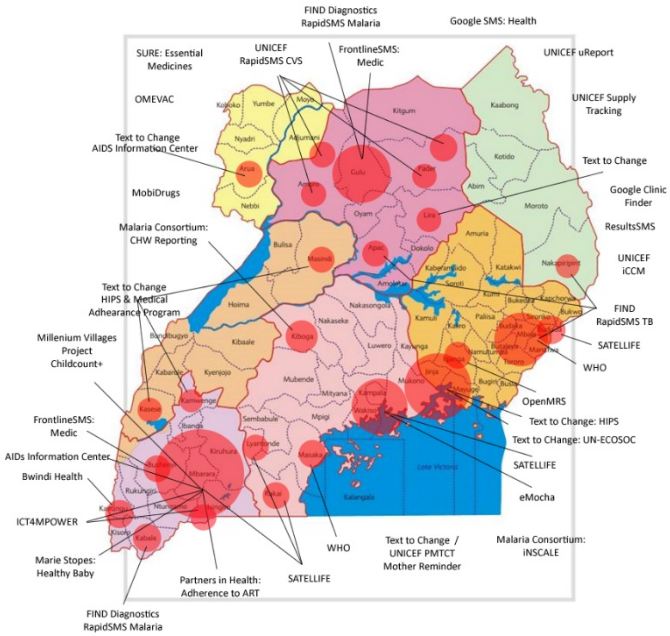


Figure 11: Uganda pilotitis (Sean Blaschke), 2011

Study 2: How should researchers evaluate "success" for mHealth implementations?

- The mHealth solution from Study 1 was taken to Malawi, but analysis of infrastructure showed that different locales had differing needs
- Mobile phones taken from India that were 'supposed to work', did not work because existing infrastructure didn't have network configuration for those devices.
- **Meeting local needs means "success" in mHealth implementations**
- Particularly important since mobiles have deeper penetration in the social fabric and blurred lines between personal and work. **Understanding praxis needs deeper research**.
- Mobility of time-space changes the idea of context. Infrastructure needs to be broken down in locales – a combination of local architecture, governance and praxis.
- Scaling as a parameter of "success" is granular. And meeting local needs in each locale becomes the recipe of scaling.


COLD CHAIN INVENTORY
CODE : LL009

Study 3: What software development methodology can be used to increase user participation in open-source communities?

- To understand local needs, one needs *translation* done by *gateways* and *boundary objects*. User participation is a gateway to translate this in open-source communities
- **The OpenScrum agile software development methodology improves developer as well as user participation in FOSS communities.**
- The study concludes that OpenScrum, a tweaked agile methodology has empirical basis by which it has helped improve bus-factor in the OpenMRS community
- Agility has been described as a strategy to adapt to changing needs. But how does one get to truly understand “changing needs” in domain-heavy infrastructure.

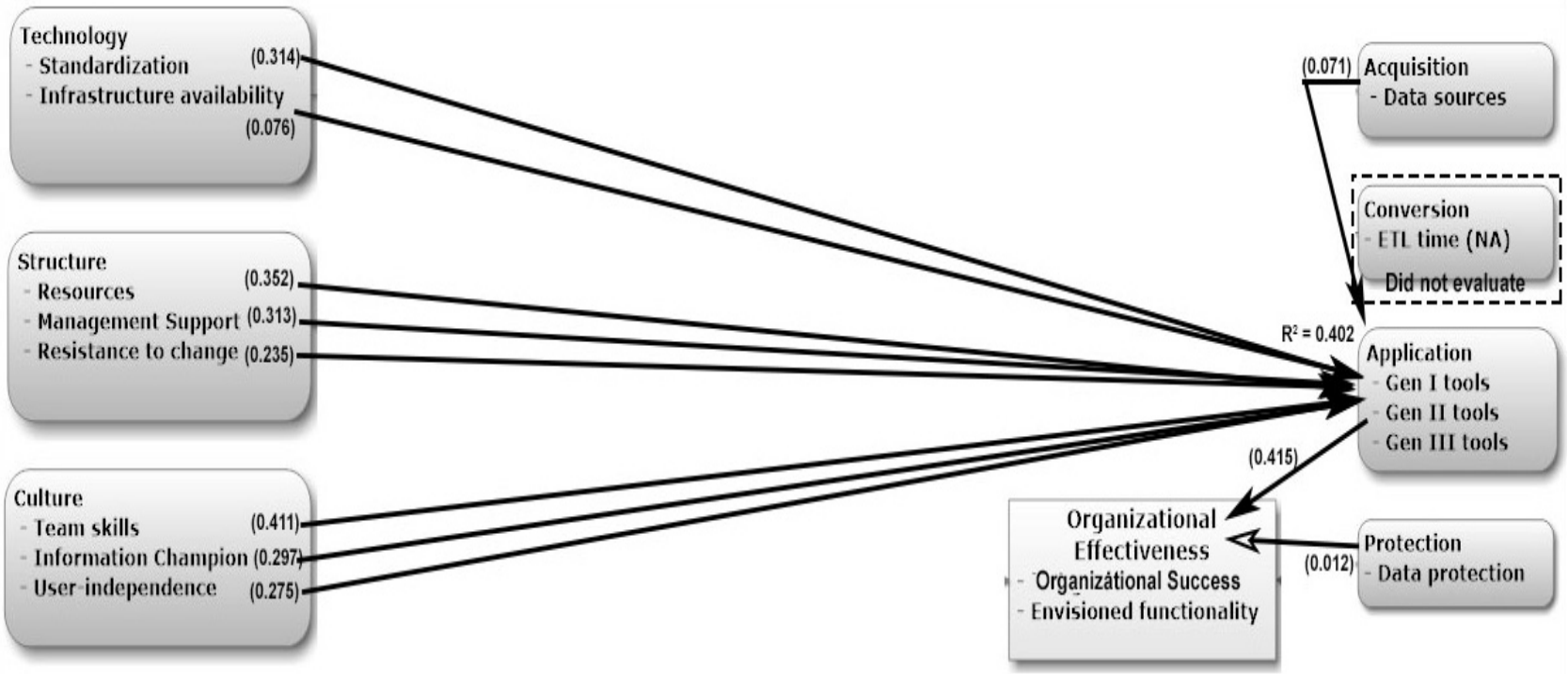
Study 4: How do security challenges arise due contextual software development?

- Contextualization of software, using the concepts of fit, alignment or meeting user needs is important in IS literature
- The move from non-networked to networked world, involves nuance to process contextualization. Without this nuance, security problems arise in the infrastructure.
- **Contextual development *inscribes* security problems in software where contexts of use are insecure.**
- The study particularly challenges the current practice in security certification process, where reasons for design choices are rarely acknowledged and never integrated as part of the testing process.
- The study makes a practical case about the limitations of understanding contextualization. It is beginning of a more critical framework in information security to better study the inscription of behavior in the infrastructure.

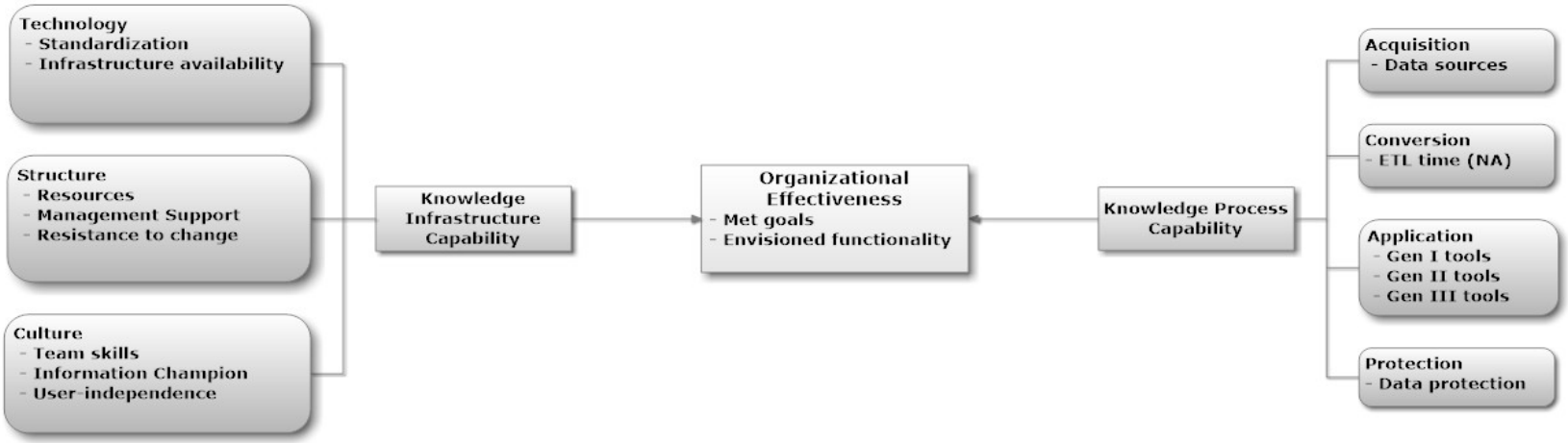
Study 5: How do warehousing and BI tools evolve to make a health system use big data?

- The study was conducted in two phases.
  - Survey in different countries to create a correlation between the factors that determined successful use of BI tools and the organizational capabilities to use those tools.
  - Highlights the evolution of the different BI tools in DHIS2.
- Using PLS to compare the variables and draw correlations between the factors of organizational capabilities and data warehouse success.
- **3 generations of Operational BI tools and Big Data is dependent on Organizational Capabilities.**
- With each generation of BI Tools in DHIS2, the tools were created to match the capabilities of the health system in low-resource settings.
- With the spread of internet and larger datasets, the Architecture of DHIS2 had to evolve to support the infrastructure.

Study 5: PLS results to define formative and reflective factors



Study 5 and 6: Organizational Capabilities




dhis2
Data Set
Data Set Section
Data Set Assignment Editor
Data set management

Study 6: What are new cloud computing models that can reduce digital divide in LMICs for analytics in IeHI?

	IT adoption stages				
Individual Organization Global	Digital access divide	⇨	Digital capability divide	⇨	Digital outcome divide

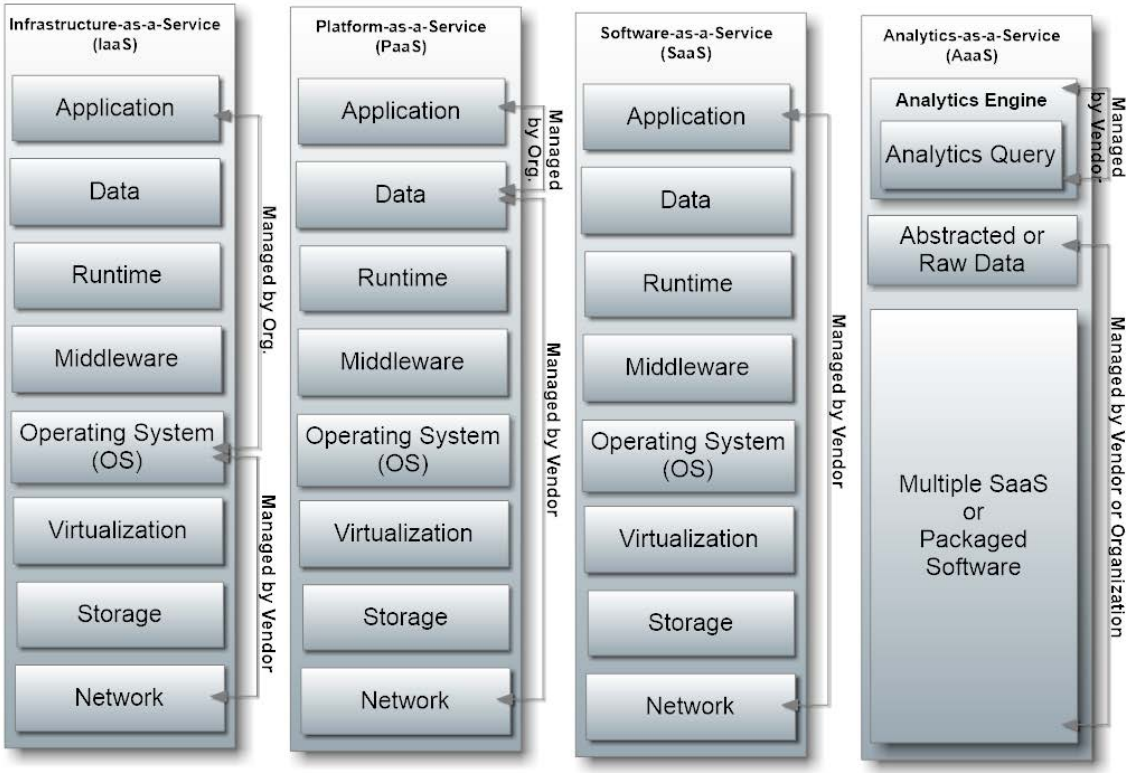


Table 8: Analysing cases through the AGP model

Cases	Architecture (A)	A-G	Governance (G)	G-P	Praxis (P)	A-P
Study 1: Design and implementation of mobile-based technology in strengthening health information systems: aligning mHealth to infrastructures (Purkayastha, 2012)	flexible/enabling architecture - open architecture - shared resources of using the DHIS2 infrastructure to report data	- Limited causality with governance activities	- enabling and participative governance regime - Users are encouraged to create their own applications and customizations to the mobile devices	- peer-to-peer communication enabled through CUG plans - capture daily data from health workers (big data) - group communication over SMS using special codes in the Kenyan case	- users expand new datasets, use-cases on the mobile reporting and SMS application - open-source technology allows modification to code base and different freedoms to innovate	- Limited causality between architecture and praxis
Study 2: Post-development perspective on mHealth – an implementation initiative in Malawi (Purkayastha, Manda and Sanner, 2013)	- a mix of JavaME, browser based application for reporting - open architecture where devices with different capabilities can join	- Billing plans for data services to monitor the differences between users of JavaME application and browser	- decentralized decision making on use of mobile plans, devices and reporting application - centralized feedback mechanism to compare facilities on health indicators	- decentralization results in improved articulation of user needs - praxis evolved its own ways to report to the health system and improve reporting rates	Organizational efficiency, such as reporting rate, cost savings, feedback were better received than planned - measuring health outcomes became secondary to project evaluation	- change from single JavaME app to a mixed architecture of browser and multiple mobile providers - a device agnostic platform for data reporting evolved instead of initial plan of JavaME app

Table 8: Analysing cases through the AGP model

Cases	Architecture (A)	A-G	Governance (G)	G-P	Praxis (P)	A-P
Study 3: OpenScrum-Scrum methodology to increase bus factor in an open-source community (Purkayastha, 2015)	- increased focus on modularity - loose coupling between core and modules	Limited causality between architecture and governance activities	- a tweaked methodology to work in sprints - community participates on same codebase for short bursts of time - Sprints, spikes, code reviews, standup meetings improve knowledge sharing in the community	- developers feel more empowered and understand codebase better - community feels greater ownership of the codebase	- sprints decrease productivity of top developers, but brings common productivity among developers - users participate more often as product owners, and developers have better understanding of codebase - less agile, but better understanding of needs	- application releases focuses on more modularity - releases of modules speeds up compared to slowness

Table 8: Analysing cases through the AGP model

Cases	Architecture (A)	A-G	Governance (G)	G-P	Praxis (P)	A-P
Study 4: Towards a contextual insecurity framework: How contextual development leads to security problems in information systems (Purkayastha, 2010)	- centralized architecture of sharing patient records and aggregate data - aggregate data is architecturally separate from patient-level data	- Limited causality between architecture and governance activities	- distributed software development - global team does not closely monitor activities for local security certification	- Limited causality between governance and praxis	- in paper systems patient record is available on at facility - passwords are shared between users due to limited internet connectivity - users lock themselves out with repeated incorrect passwords when they do not want to work	- password reset not available with user and needs admin intervention - patient records are available centrally to all health facilities - encryption not required for aggregate statistics, but essential for patient data that is centrally shared

NTNU

Table 8: Analysing cases through the AGP model

Cases	Architecture (A)	A-G	Governance (G)	G-P	Praxis (P)	A-P
Study 5: Overview, not overwhelm: Operational BI tools 1 for Big Data in health information systems (Purkayastha & Braa 2015)	- internet based tools from offline Access software - high performance analytics tools that are flexible - 3 generations of BI tools are based on shared and open architecture	- centralized data storage and processing - core team that globally determines needs and flexible architecture	- hierarchy of standards based on what is required at each level of the health system - decentralized decision making process - BI tools allow mapping and comparing health facilities	- Limited causality between governance and praxis	- organizational structure and culture determine success of BI tools - even with availability of technology, limited use due to structure and culture factors - low sustainability of technical skills	- Limited causality between architecture and praxis

Table 8: Analysing cases through the AGP model

Cases	Architecture (A)	A-G	Governance (G)	G-P	Praxis (P)	A-P
Study 6: Big Data Analytics for developing countries: Using the Cloud for Operational BI in Health (Purkayastha & Braa 2014)	- new scheduler for analytics based on computing resource availability - open architecture that can scale on multiple servers	- new architecture for deployments enable use of cloud computing providers - resource sharing between countries possible due to shared infrastructure	- move IS management to external vendor - use cloud computing models of IaaS, PaaS, SaaS or AaaS for using external services	- integrate capabilities of external vendors into the infrastructure	- Limited changes to praxis due to new business models	- Limited causality between architecture and praxis

Blind spots in the AGP model

- Visibility of work is often a challenge for researchers to study, as has been well documented in the CSCW field
- **Blind spot to micro-level methodologies and concepts** – This is particularly seen in Study 2 and Study 3 where they focus on methodological or theoretical notions, but the AGP model doesn't seem to clearly articulate their necessity
- **Blind spot to notions beyond work** – There is a whole body of concepts for example structure, culture, politics, counter-networks that are not well understood as work.
- **Blind spot to external activities of the infrastructure** – This comes from the fact that infrastructure is not everything of everything. All agencies that act forces of causality on components of the infrastructure cannot be considered part of the infrastructure. There are boundaries that help articulate what work is done within and for the establishing of the infrastructure

Conclusion

- The AGP model is a taxonomy that provides an organizing thought to infrastructuring work
- The AGP model is a holistic view to the different categories of work
- The AGP model can be applied post-facto to study the evolution of an infrastructure or as a way to organize teams and their work in the activities of establishing an infrastructure
- The AGP model is still a macro-level view and hence is described as a genus-level taxonomy. There are higher and lower level taxonomies which could describe the work within architecture, governance and praxis categories

Future work

- Health information exchanges in the US. Expanding on the infrastructure work to describe Indiana Health Information Exchange
- Describe the capabilities from different participating organizations and clinicians in
- Using repotoire grids to analyze capabilties and self-suffciency to describes lower-level taxonomic ranks of activities in an infrastructure

To the opponents!!