

Submission No: 10932

PDW Proposal for the 2011 Academy of Management Conference

Submission No: **10932**

TITLE:

Decoding the DNA in social sciences: a sequence analysis method for analyzing complex process data

PRIMARY SPONSOR: OCIS

CO-SPONSORS: RM, ODC, OM, OB, TIM

ABSTRACT

Until recently, the use of sequence analysis techniques in social studies have been primarily limited to analyzing the temporal order and pattern of critical activities (Abbott, 1990; Shoval & Isaacson, 2007). Yet, the method has much more to offer – especially for IS and organizational process scholars who are interested in evolutionary patterns of events over time. We offer an extended method for analyzing not only the temporal sequence of activities, but also variation in their constituent parts, the patterns within and between them, and their mutation over time. This new approach is an extended version of the event sequencing approach introduced by Gaskin et al (2010a). In this five-hour workshop we will discuss the key elements of this approach and demonstrate uses of this approach in a hands-on format across a number of application areas. Thus we encourage participants to bring their own process data that could be subjected to sequence analysis.

This PDW will consist of two main parts. The first part will be used to present the key ideas and the basic steps of the method. The second part will be used to assist participants in discovering how they can apply sequence analysis to their own data.

The main goals of this PDW are to: 1) provide scholars with a useful new method for analysing complex social data; (2) generate discussion and feedback to help improve this novel method; (3) increase awareness of other researchers' work in this field; and (4) encourage collaboration and new research ideas.

PDW OVERVIEW

The sequences of human activities, such as work processes, buying behaviors, or communication patterns can be analyzed using space and time series (Shoval & Isaacson, 2007), where space and time are used as reference points for estimating the sequences of events. We have been working on an extension of this method which enables us to detect changes in the order and configuration of any human activity if it can be modeled as a variation over a set of attributes across some order (typically temporal) (Gaskin et al., 2010a). Accordingly, this sequence analysis approach can be expected to reveal underlying structure of the activities (i.e. how activities or events relate to one another, and how they vary from one activity to the next while varying some base elements) (Gaskin, Schutz, Berente, Lyytinen, & Yoo, 2010b). Using this technique we can also detect underlying generative mechanisms including a range of mutations and the evolution in the “DNA” of complex social activity as we probe its structure over time. This allows us to explore what new elements are introduced or how some basic elements are sequenced differently.

Previously, sequence analysis has been used in social studies to analyze variance in spatial or temporal behaviors (Shoval & Isaacson, 2007; Wilson, 2001, 2006) and to reveal patterns of social change (Abbott, 1990). In the IS field, Pentland (2003) used a similar technique to measure variations in administrative work processes, Sabherwal and Robey (1993) to study variation in IS implementation processes, and Sabherwal and Grover (2010) to study variations in political tactics during implementation. However, none of these analyses have been extended to explicitly tap into and capture variations and patterns generated by different configurations of chosen set of elements within social activities.

In Gaskin et al. (2010a) we proposed a method that can partially address such needs. In the workshop we will discuss how to represent and analyze patterns of activities and how they evolve. Specifically, we will use examples from a set of design processes that have been

drawn from complex real life settings in large design processes carried out by different organizations. Using this method, we show how it is possible to identify patterns in the design routines at different levels that make up the design routine (actors, location, tools, type of activity, etc.).

This method is now being used to identify patterns in research areas as diverse as IT adoption behaviors and open source discussion forums in capturing and analyzing discussion threads. The IT adoption patterns is expected to cluster classes of organizations based on the observed impact of adopting a certain innovation outcome and hence to evaluate the innovation potential of different technological choices. The forum communication patterns are expected to determine clusters in forum responses that can predict successful resolution or failure in resolving problems.

Our main idea is to show that the method can be used for different types of complex socio-technical data where we assume that there are certain underlying structural relationships among data elements. While this method is still very much in a formative stage of development, we have already seen its potential, with regards to analyzing complex social data sets, to:

- Identify patterns
- Reveal evidence of mutations and evolution
- Identify clusters/classes/categories of behavior or structure
- Reveal intricate low-level details hitherto difficult to discover in a mass of data

One of our aims for this workshop is to identify other contexts for use and discuss new analysis capabilities that could inform IS or technology research.

The PDW consists of two parts: a three-hour afternoon session followed by a two-hour session. Both sessions are open to all registered participants (maximum = 50). The first session will be primarily informative and instructive. The second session will be used to help participating scholars use the method with their own data or research problem. Those

without their own data are free to stay and observe or leave at any time during the second session.

**THE FIRST SESSION (3 hours):
Sequence Analysis Method**

The first session is devoted to discussing the state of the art with regards to sequence analysis, and then to learn hands-on, how to perform a sequence analysis with complex social data. This session will be structured as follows.

1) The state of the art in Sequence Analysis (60 minutes):

Speaker: Kalle Lyytinen. The DNA of Design: using sequence analysis to analyze design processes.

This part will present materials outlined in Gaskin et al. (2010) “Sequencing Design DNA: A Set of Methodological Artifacts for Sequencing Socio-technical Design Routines” including new developments that have taken place after that paper’s publication. We will also offer examples of other uses of sequence analysis method in analyzing data. (30 minutes including Q&A).

Speaker: Youngjin Yoo, Why do we care? The Evolutionary Sciences of the Artificial

In this talk, we will discuss the potential emergence of Evolutionary Sciences of the Artificial as a new scientific discipline and organizational genetics as a particular branch that is concerned with the study of evolution of man-made artifacts including organizations and technology. We will discuss other potential applications of this approach in a broad scientific context. (30 minutes including Q&A).

~Brief Break~

2) Hands-on Tutorial (1.5 Hours):

During this portion of the session we will show participants the tools used to analyze a sample dataset of design processes. Participants desiring to follow along during this portion

will need to have the following software installed and running properly prior to the workshop (we cannot take time during the workshop to work out these types of issues):

- Excel 2007 or 2010 for Windows (2008 for Mac WILL NOT WORK). Microsoft Office 2010 can be downloaded as a trial version from <http://office.microsoft.com/en-us/try/>. HOWEVER, THIS TRIAL VERSION WILL REMOVE PRIOR VERSIONS, so do not install if you have prior versions and you do not intend to actually buy the 2010 version after the trial expires.
- MetaEdit+. Trial version available from <http://www.metacase.com/download/>
- ClustalG. Full version publically available from www.kolobkreations.com/clustalg1_5.zip
- MEGA. Full beta version available publically from <http://www.megasoftware.net/beta/index.php>

During this session we will model a design scenario in MetaEdit+, then scrape it for information using the code in appendix 1. We will also discuss how to modify and extend the types of data that can be captured by the MetaEdit+ tool. We will then restructure that visual process data in Excel, perform sequence alignment in ClustalG, and then perform cluster analysis in MEGA (using the sequences in appendix 2). We can then explore lower level details of the data using Excel's pivot tables. All this will be done to show new patterns and clusters. Lastly, we will perform a similar analysis to show the *evolution* of design processes using the sequences in appendix 3.

(90 minutes)

~Brief Break~

THE SECOND SESSION (2 hours)

Decoding Your Own Data and Planning Future Research

The second session has two main aims: 1) applying this method to participants' own datasets and 2) networking to plan for future research collaborations. Participants who come with their own data will receive one-on-one assistance from the four facilitators (James Gaskin, Omri Shiv, Veeresh Thummadi, and Yohan Choi). We will assist participants in discovering how their own data can be explored for patterns, clusters, mutations, and evolution.

Participants who do not bring their own data can observe or listen-in as others explore and discover. Hopefully, observing will provide insight into how their data can be analyzed with this new method. As the session progresses, we hope that participants will branch out and seek opportunities to collaborate and plan new research.

REGISTRATION

Registration for both sessions is optional, and those who have not registered will not be turned away unless there are not enough seats for those who have registered. However, registration is encouraged for two reasons: first, because this will give us a better idea of how many participants to plan for, and second, because we will be able to email detailed instructions regarding software and data prior to the workshop.

RELEVANCE TO OCIS:

The aims of this PDW will assist scholars that have the same interests sponsored in OCIS to analyze their data in ways previously unfeasible. Many OCIS scholars have complex social data that could previously only be analyzed through content or thematic analyses. This PDW will offer them another option which will also provide a new perspective on old data.

RELEVANCE TO OTHER AOM DIVISIONS

We believe that this PDW will be of interest to several AOM divisions, for example:

Relevance RM:

As we are introducing a new research method, this is highly relevant to RM. We hope that this new method will attract new attention and scrutiny, in order to improve the method for all who utilize it.

Relevance ODC:

The data that we will be analyzing during the PDW is a set of organizational routines that change over time. This kind of data is particularly well-suited to this new method.

Relevance to OM:

Managing operations includes understanding activity sequences and patterns – including potential areas for change. This new method makes that kind of analysis possible.

Relevance to OB:

Organizational routines and behavioral patterns are complex social data that we can decode using this new method. This method will provide researchers in OB a new vantage point on their data.

Relevance to TIM:

Part of the data we will be analyzing during this PDW is design processes in virtual building and construction. We will be exploring the role and impact of technology in the evolution of design processes.

PDW ORGANIZERS	
James Gaskin	Case Western Reserve University Cleveland OH james.eric.gaskin@gmail.com 216-245-7050
Kalle Lyytinen	Case Western Reserve University (USA)
Youngjin Yoo,	Temple University (USA)
OTHER PDW CONTRIBUTORS	
Veeresh Thummadi	Case Western Reserve University
Omri Shiv	Case Western Reserve University
Yohan Choi	Temple University (USA)

STATEMENT OF PARTICIPATION

I have received signed statements from all intended participants agreeing to participate for the entire workshop, AND stating that they are not in violation of the Rule of Three + Three.

James Gaskin

REFERENCES

- Abbott, A. 1990. A primer on sequence methods. *Organization Science*: 375-392.
- Gaskin, J., Lyytinen, K., Thummadi, V., Schutz, D., Yoo, Y., Weiss, A., & Berente, N. 2010a. Design DNA Sequencing: A Set of Methodological Artifacts for Sequencing Socio-technical Design Routines, *ICIS*. St. Louis, MO.
- Gaskin, J., Schutz, D., Berente, N., Lyytinen, K., & Yoo, Y. 2010b. The DNA of design work: Physical and digital Materiality in project-based design organizations, *Academy of Management Best Paper Proceedings*. Montreal, Canada.
- Pentland, B. T. 2003. Sequential Variety in Work Processes. *Organization Science*, 14(5): 528-540.
- Sabherwal, R., & Grover, V. 2010. A taxonomy of political processes in systems development. *Information Systems Journal*, 20(5): 419-447.
- Sabherwal, R., & Robey, D. 1993. An empirical taxonomy of implementation processes based on sequences of events in information system development. *Organization Science*, 4(4): 548-576.
- Shoval, N., & Isaacson, M. 2007. Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American Geographers*, 97(2): 282-297.
- Wilson, C. 2001. Activity patterns of Canadian women: Application of ClustalG sequence alignment software. *Transportation Research Record: Journal of the Transportation Research Board*, 1777(-1): 55-67.
- Wilson, C. 2006. Reliability of sequence-alignment analysis of social processes: Monte Carlo tests of ClustalG software. *Environment and Planning A*, 38(1): 187.

Appendix 1 – MetaEdit+ scraping code

```

Report 'EssentialsOnly'
foreach .Activity
{
/* This gets the properties of Activity*/
id ' '
do :()
{ if id;l <> id then
id ' '
endif
}
/**END activity PROPERTIES**/
/* This gets the connected Actor and it's properties*/
dowhile >Actor-activity
{
do .Actor
{
id ' '
do :()
{ if id;l <> id then
id ' '
endif
}
}
}
/**END Actor **/
/* This gets the connected Tools and their properties*/
dowhile >Activity-Tool
{
:Affordance type;
' '
do .Tool
{
id ' '
do :()
{ if id;l <> id then
id ' '
endif
}
}
do ~From tool { :Data flow; ' ' }
/* This gets the connected DO and it's properties*/
dowhile >Tool-Design object
{
do .Design Object
{
id ' '

do :()
{ if id;l <> id then
id ' '

endif
}
}
}
/**END DO**/
}

```

```
}  
}  
/**END Tool**/  
/**END Activity**/  
newline  
}  
endreport
```

Appendix 2 – Sequence Data

```
>s1_Upload_Issued_for_Construction_Design_Doc  
TfrRdtGrpStoDigOutOutProSpeSpe  
>s2_Obtain_Issued_for_Construction_Design_Doc  
ValRdtGrpAnaDigInpInpProSpeSpe  
>s3_Develop_Separate_trade_3D_models  
GenRdtGrpTraDigInpPro  
>s4_Upload_Separate_trade_3D_models  
TfrRdtGrpStoDigOutPro  
>s5_Generate_composite_3D_model  
GenColGrpConDigInpPro  
>s6_Generate_Clash_Report  
GenColGrpCopDigInpInpPrcPro  
>s7_Upload_composite_model_and_clash_reports  
TfrRdtGrpStoDigOutOutPrcPro  
>s8_MEP_Coordination_Meeting  
NegColGrpAnaDigInpInpPrcPro  
>s9_Separately_resolve_clashes_update_and_repost  
NegRdtGpxCopStoTraDigDigDigInpOutOutOutInpPrcProPro  
>s10_Generate_composite_3D_model  
GenColGrpConDigInpPro  
>s11_Generate_clash_report  
GenColGrpCopDigInpInpPrcPro  
>s12_MEP_Coordination_Meeting  
NegColGrpAnaDigInpInpPrcPro  
>s13_Create_sign-off_documents  
GenColGrpConDigOutOutOutImpImpImp  
>s14_Fabrication_of_systems_per_sign-off_models  
GenRdtIngTraPhyInpInpInpSpeImplImp  
>s15_Post_Drawings_CD/ASI/COB  
TfrRclGrpStoDigOutPro  
>s16_Update_M/P/FP/E_engineering_systems  
GenRdtGpxTraDigUpdUpdSpePro  
>s17_Post_Architectural_&_Construction_models  
TfrRclGrpStoDigOutPro  
>s18_3D_MEP_coordination_on_M/P/FP/E_systems  
GenRdtGpxAnaDigInpOutProSpe  
>s19_Post_M/P/FP/E_coordinated_3D_models  
TfrRdtGpxStoDigOutPro  
>s20_Generate_Clash_Report  
GenColGrpAnaDigOutSpe  
>s21_Review_models  
ValColGrpRepDigInpInpInpProSpePro  
>s22_MEP_Coordination_Meeting  
NegColGpxCopRepRepRepDigDigDigPhyOutInpInpOutOutOutOutSpeProProSpe  
>s23_Post_3D_concrete_penetrations  
TfrRdtGpxStoDigOutSpe  
>s24_McKinstry_&_Cochran_resolve_conflicts
```

GenRdtGpxAnaDigInpUpdUpdSpeProPro
>s25_Generate_clash_reports
GenColGrpAnaDigOutSpe
>s26_Sign_off
ValColGrpRepDigOutPrc
>s27_Post_Drawings
TfrColGrpStoDigInpPro
>s28_Update_Construction_Model
GenDstInxAnaTraDigDigInpOutInpOutProPro
>s29_Post_3D_Models
TfrColIndStoDigInpPro
>s30_MEP_Coordination
NegRdtGpxStoTraDigDigInpUpdPro
>s31_Post_Coordinated_Model
TfrRdtGpxStoDigInpPro
>s32_Review_Coordinated_Model
ValColIndAnaDigInpPro
>s33_MEP_Coordination_Meeting
NegColGpxAnaCopSupDigDigPhyInpInpInpOutOutProSpePrc
>s34_Approve_Submittals
ValColIndAnaStoDigPhyInpOutImpPrc
>s35_Generate_CP_Drawings
GenDstGpxCopDigInpSpe
>s36_Generate_Clash_Report
GenColIndAnaDigInpOutImpSpe
>s37_Sign_Off
ValColIndCopDigOutPrc

Appendix 3 – Evolution Data

>2007
TfrRdtGrpStoDigOutOutOutProSpeSpeValRdtGrpAnaDigInpInpInpProSpeSpeGenRdtGrpTraDigInpProTfrRdt
GrpStoDigOutProGenColGrpConDigInpProGenColGrpCopDigInpInpPrcProTfrRdtGrpStoDigOutOutPrcProNe
gColGrpAnaDigInpInpPrcProNegRdtGpxCopStoTraDigDigDigInpOutOutOutInpPrcProProGenColGrpConDig
InpProGenColGrpCopDigInpInpPrcProNegColGrpAnaDigInpInpPrcProGenColGrpConDigOutOutOutImpl
mpGenRdtIngTraPhyInpInpInpSpeImpImp
>2008
TfrRclGrpStoDigOutProGenRdtGpxTraDigUpdUpdSpeProTfrRclGrpStoDigOutProGenRdtGpxAnaDigInpOut
ProSpeTfrRdtGpxStoDigOutProGenColGrpAnaDigOutSpeValColGrpRepDigInpInpInpProSpeProNegColGpx
CopRepRepRepDigDigDigPhyOutInpInpOutOutOutOutSpeProProSpeTfrRdtGpxStoDigOutSpeGenRdtGp
xAnaDigInpUpdUpdSpeProProGenColGrpAnaDigOutSpeValColGrpRepDigOutPrc
>2009
TfrColGrpStoDigInpProGenDstInxAnaTraDigDigInpOutInpOutProProTfrColIndStoDigInpProNegRdtGpxSto
TraDigDigInpUpdProTfrRdtGpxStoDigInpProValColIndAnaDigInpProNegColGpxAnaCopSupDigDigPhyInp
InpOutOutProSpePrcValColIndAnaStoDigPhyInpOutImpPrcGenDstGpxCopDigInpSpeGenColIndAnaDigIn
pOutImpSpeValColIndCopDigOutPrc

1. Welcome – James (Introduce Workshop, participants, agenda)
2. Post <http://kolobkreations.com/aompdw.html>
3. Explain background and motivation for developing this kind of a method (if Kalle and Youngjin have not already done so)
4. Demo the method and follow along if desired.
 - a. Explain MetaEdit
 - b. Build a small model
 - c. Show completed model (AOM PDW)
 - d. Extract data, Restructure
 - e. Transition matrices
 - f. Pivot Table
 - g. ClustalG
 - h. Mega
5. Break and then work one on one.