

Risk Analysis Tool Setup Manual

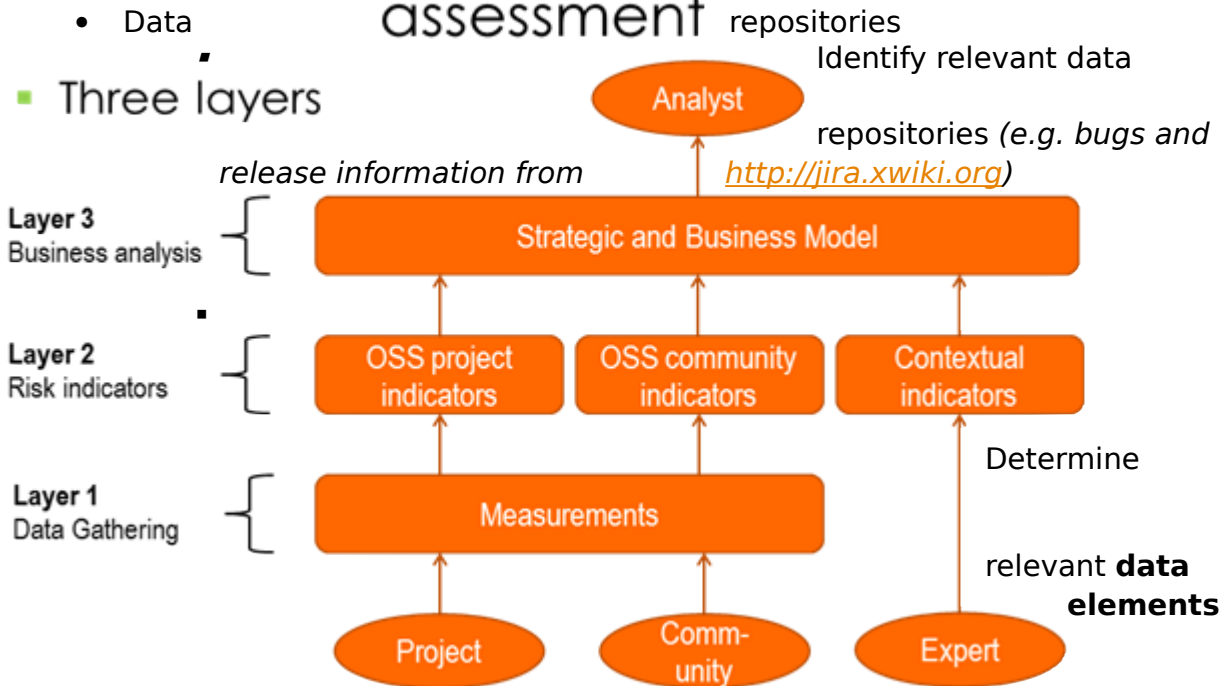
Submission Date:	1.13.2015	
Version:	<1.0>	
Status:	<draft final>	
Author(s):	R. Ben-Jacob	KPA
	Ron Kenett	KPA
Reviewer(s):		

1 Introduction

This document presents technical details of the setup required to deploy a Bayesian network risk model to be incorporated on the RISCOSS Platform. The purpose of the document is to provide implementation steps for data preparation and creation of the Bayesian networks which will comprise a risk model. The tools invoked in the setup process are also documented and the terminology used is in line with the RISCOSS general ontology.

The graphical tool used to create the Bayesian networks in xdsl format GeNIe (Graphical Network Interface). This tool is the graphical interface to SMILE (Structural Modelling, Inference, and Learning Engine), a fully portable Bayesian inference engine developed by the Decision Systems Laboratory and thoroughly field tested since 1998. GeNIe can be freely downloaded from <http://genie.sis.pitt.edu> with a user guide and related documentation.

2 A layered approach for risk assessment Data Sources



© RISCOSS Consortium

First Year Evaluation, January 2014, Brussels

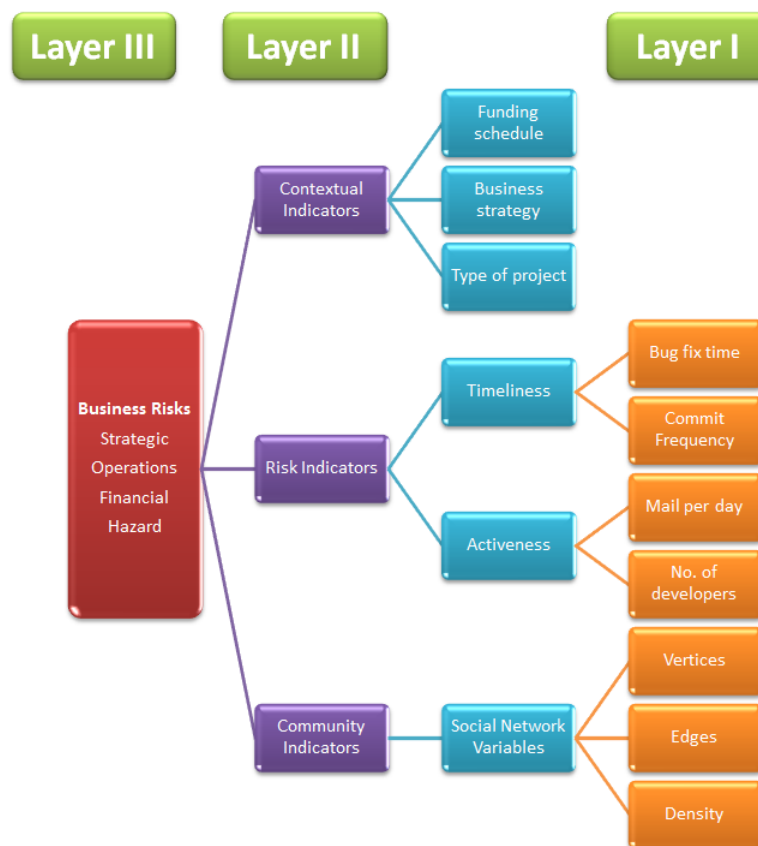
6

(**measurements**) for each data repository (e.g. 'bug fix time' or 'Number of Unique Licenses')

- Initial retrieval of all data elements from all relevant data sources
- Additional information
 - Social Network Analysis (SNA)
 - Determine relevant **SNA metrics** (e.g. 'Density')
 - Obtain initial data from email or chat logs (e.g. IRC chat archive from <http://dev.xwiki.org/xwiki/bin/view/IRC/WebHome>)
 - Perform social network analysis (e.g. NodeXL <https://nodexl.codeplex.com/>)
 - Additional information
 - Determine required **contextual indicators** (e.g. 'Business Strategy')
 - Define possible values for each contextual information
 - Prepare online questionnaires

3 Define risk model structure

- Risk models are based on three-layer approach:
 - Layer-1 **Risk Drivers**: the distribution levels of each data element or SNA metrics (e.g. 'Commit Frequency' or 'Licenses and IP')
 - Layer-2 **Risk Indicators**: the aggregated effect of related risk drivers (e.g. 'Activeness' or 'Quality')
 - Layer-3 **Business Risks**: the result effect of all risk indicators and contextual indicators (e.g. 'Strategic')



- Create risk drivers: for each data element and SNA metrics, define its distinct **distribution groups**. Use any of the following methods:
 - Statistically: basic (mean, median, standard deviation) or advanced statistics, or use of Bayesian discretization function (see **Genie** <https://dslpitt.org/genie/>)
 - Expert assessment based on knowledge and experience

The result of this step is illustrated thusly: data element 'Mail Per Day' is grouped into 5 distribution levels: 0, 1 to 7, 8 to 25, 26 to 100, 101 and up

2. Define risk indicators: identify all risk drivers that affect a risk indicator.
Note: typically all SNA metrics are aggregated into one risk indicator (e.g. 'Community Indicator')
3. Define number of levels for risk indicators (i.e. low, medium, high)
4. Define business risks
5. Define number of levels for business risks (i.e. very low, low, medium high, very high)
6. Note: contextual indicators directly affect business risks. They do not have distribution level per se since they can be set to one value at a time

4 Create layer-2 scenarios

- The basis for the creation of indicator **Bayesian networks** (BN)
- Initially, create about 50 scenarios per risk indicator
- Scenario is a collection of values of all data elements per risk indicator. They can be either:
 - Real data scenarios: built based on real data (e.g. OW2 EasyBeans or Joram), or
 - Simulated scenarios: randomly built based on range of possible values per data element
- Risk driver and risk indicator names: use only letters and numbers, substitute spaces with an underscore (_)

5 Risk Indicator levels

- For each layer-2 scenario, determine its outcome risk indicator value
- Use any of the following methods:
 - **Tactical Workshop**: several experts assess indicator level for each scenario
 - **Calculate**: based on expert assignment of weights for each distribution group (i.e. its relative impact on the risk indicator)

6 Create layer-3 scenarios

- The basis for the creation of business risk Bayesian network
- Initially, create about 50 scenarios
- Scenario is a collection of simulated values of all risk indicators and contextual indicators
- Business risks and contextual names: use only letters and numbers, substitute spaces with an underscore (_)
- Risk indicator names should be identical to names used for layer-2

7 Business Risks level

- For each layer-3 scenario, determine its outcome business risk indicator
- **Strategic Workshop:** several experts assess indicator level for each scenario

8 Create risk indicator Bayesian networks, using GeNIe

- Discretize input data elements based on previously determined distribution groups. Note: make sure all levels exists (e.g. use *Uniform method and manually adjust to-from values*)
- Discretize output risk indicator based on previously determined levels. Note: make sure all levels exist
- Create simple network: arcs should to go only from input risk drivers to output risk indicator
- Large output **xdsl** files (*consult system administrator*) can be made smaller by creating a 'reduced model' by removing nodes based on one of the following analyses:
 - 'Sensitivity analysis' with 'target' risk indicator
 - 'Influence' analysis, with one of the available strength/distance combinations
- Set 'User Property'
 - Network: Type=INDICATOR
 - Input risk driver nodes: Type=INPUT
 - Output risk indicator: Type=INDICATOR

9 Create business risk Bayesian networks

- Discretize input risk indicators and contextual indicators based on previously determined distribution groups. Note: make sure all levels
- Discretize output business risks based on previously determined levels. Note: make sure all levels exist
- Create simple network: arcs should to go only from input risk and contextual indicators to output business risks
- Set 'User Property'
 - Network: Type=RISK
 - Input contextual indicator nodes: Type=INPUT
 - Input risk indicator: Type=INDICATOR
 - Output business risks: Type=RISK
- Note: Risk indicator node must have the same **node id** on indicator and risk BN's

10 Deployment

- All out xdsl files should be placed in the same folder
- The analysis tool loads all BN files and requests values for all Type=INPUT nodes (risk drivers and contextual indicators)
- Values should be provided as list of distribution levels, totaling 1 (100%) per risk driver or contextual indicator
- Values are applied to the Bayesian networks as 'virtual evidence'
- The analysis process is performed following these steps:
 - Obtain input risk driver and contextual indicator values from the RISCOSS platform (which will run the necessary data collectors)
 - Apply risk driver values to indicator networks
 - Obtain interim output risk indicator distribution levels
 - Apply input contextual indicator and interim risk indicator levels to business risk network
 - Obtain output business risk levels
 - Return risk indicator and business risk levels for rendering by the RISCOSS platform

