

CONI  
SOFT  
2020

# 2020 The 8th International Conference in Software Engineering Research and Innovation

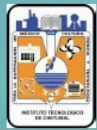
Tecnológico Nacional de México  
Instituto Tecnológico de Chetumal



MOSCAF

## Specifying Data Quality Requirements according Web Functionalities

César Guerra-García, Hector G. Perez-Gonzalez, Marco Ramírez-Torres, Reyes Juárez-Ramírez, **Hugo Gonzalez**



# Agenda

- Introduction
- Fundamentals
- MOSCAF framework
- Results
- Conclusions

# Introduction

- Data is important these days. Data Quality ensures better performance.
- Web applications are everywhere, as an important source of data.
- Potholes can exist in these applications.

- Our research goal is to provide developers with an adequate methodological **mechanism and artifacts** to manage **DQ requirements** within **Web application development**.
- Our concern was to explore how to capture specific data quality requirements.

- We designed a working strategy whose ultimate goal was to obtain a **generic model** (MOSCAF) to help designer to understand **DQ requirements** for specific functionalities of a web application.



# Fundamentals

DQ  
Web functionality

# Data Quality

- A measurement on “how good” is the data.
- Wang and Strong define data quality dimensions
  - Intrinsic, Accessibility, Contextual, Representational
- ISO/IEC 25012 is a generic DQ model
  - 15 characteristics from two points of view. Inherent and system-dependent

# Web functionality

- Collins describe 11 web functionalities
  - Content management, process and actions
  - Search capability, Administration, Security
  - Data points and integration, Presentation
  - Communication and collaboration, Taxonomy
  - Personalization, Help.



# Known potholes (Strong et. al)

- Multiple sources, subjective production
- Production errors, too much information
- Distributed systems, non-numeric information
- Advanced analysis requirements
- Changing task needs
- Security and privacy requirements
- Lack of computing resources

# MOSCAF: A MODEL FOR THE SELECTION OF DQ REQUIREMENTS THAT MUST BE WATCHED FOR EACH WEB FUNCTIONALITY

# Analysis process

- Web functionalities vs potholes
- DQ characteristics vs potholes
- Web functionalities vs DQ characteristics

<div>Potholes</div> <div>Web functionalities</div>	Multiple sources	Subjective production	Production errors	Too much information	Distributed systems	Nonnumeric information	Advanced analysis requirements	Changing task needs	Security and privacy requirements	Lack of computing resources
Content Management	√	√	√	√		√	√	√	√	√
Process and actions				√				√		
Search capabilities				√	√	√			√	√
Administration								√		
Security									√	
Data points and integration				√	√				√	
Communication and collaboration				√	√		√			√
Presentation								√		
Taxonomy					√					
Personalization								√	√	
Help features				√						

<div>Potholes</div> <div>DQ dimension</div>	Multiple sources	Subjective production	Production errors	Too much information	Distributed systems	Nonnumeric information	Advanced analysis requirements	Changing task needs	Security and privacy requirements	Lack of computing resources
Consistency	x									
Beliavability	x	x								
Objectivity		x								
Correctness			x							
Completeness			x					x		
Relevancy			x				x	x		
Concise representation				x		x				
Timeliness				x	x					
Value-added				x	x	x	x	x	x	x
Accessibility				x		x			x	x
Consistent representation					x		x			
Analysis requirements							x			
Security									x	

Wang & Strong model	Standard ISO/IEC 25012
Accuracy	Accuracy
Completeness	Completeness
Concise representation	Completeness
Consistent representation	Consistency
Objectivity	Consistency
Beliavability	Credibility
Reputation	Credibility
Timeliness	Currentness
Accessibility	Accessibility
Value-added	Compliance
Security	Confidentiality
Amount of information	Efficiency
Amount of information	Precision
Traceability	Traceability
Easy of understanding	Understandability
Interpretability	Understandability
Variety of data and data sources	Availability
Easy of operation	Portability
Flexibility	Portability
	Recoverability



- We rewrite the table and performed a more exhaustive analysis, concerning to other characteristics that might be simultaneously suspected to be linked with some other web functionalities
- Obtaining a set of new relationships showed with “ $\alpha$ ”.

ISO 25012 Web functionalities	Accuracy	Completeness	Consistency	Credibility	Currentness	Accessibility	Compliance	Confidentiality	Efficiency	Precision	Traceability	Understandability	Availability	Portability	Recoverability
Content Managem	$\alpha$	✓	✓	✓	✓	✓	✓	✓			$\alpha$	$\alpha$		$\alpha$	
Process and actions		✓			✓	✓	✓								
Search capabilitie		✓	✓		✓	✓	✓	✓			$\alpha$		$\alpha$		
Admon		✓					✓		$\alpha$	$\alpha$		$\alpha$		$\alpha$	$\alpha$
Security						✓	✓	✓			$\alpha$				
Data points and integraton		✓	✓		✓	✓	✓	✓					$\alpha$		
Comm and collabo	$\alpha$	✓	✓	$\alpha$	✓	✓	✓						$\alpha$		
Presentatin		✓					✓								
Taxonomy	$\alpha$		✓		✓		✓								
Personaliz		✓				✓	✓	✓					$\alpha$		
Help features		✓			✓	✓	✓					$\alpha$			

- To assure the quality of the data used through the Web functionality “Content Management”, we have to add some software artifacts aimed at implementing the following DQ characteristics of:
  - Accuracy, Completeness,
  - Consistency, Credibility,
  - Currentness, Accessibility,
  - Compliance, Confidentiality,
  - Traceability, Understandability and
  - Portability.

ISO 25012 Web functionalities	Accuracy	Completeness	Consistency	Credibility	Currentness	Accessibility	Compliance	Confidentiality	Efficiency	Precision	Traceability	Understandability	Availability	Portability	Recoverability
Content Managem	$\alpha$	$\sqrt$	$\sqrt$	$\sqrt$	$\sqrt$	$\sqrt$	$\sqrt$	$\sqrt$			$\alpha$	$\alpha$		$\alpha$	

# Conclusions

- To improve the DQ of web applications, we presented MOSCAF.
- The objective is to help on the identification of DQ requirements for a generic web application.
- We are currently in the process of validation the proposed model in a real-world web projects.



## Preguntas