

***Risk Control (Data Extraction From Endomondo)
Project Deliverable Iteration#1***

October 12, 2018

Created By:	Group Q	Date:	2018-10-12
Group Members:	Asif Butt (24532139)	Document Title:	Iteration#1 Deliverable Risk Control
	Airi Chow (40003396)	Version:	1.0
	Tung Nguyen (40042891)	Status:	Final
	Ramy Elhoufy (40001513)		

TABLE OF CONTENTS

Required Deliverables	4
Introduction	4
Purpose	4
Objective	4
Report	4
Major Data Extraction Tasks	4
Findings, Task Status & Brief Explanation	5
Data Extraction Illustration Diagram	7
Work Log	7

Required Deliverables

- Identify Information source
- Specify how you obtain the data
- Specify the data schema of the extracted data
- Provide a log file describing the contribution of each member

Introduction

Purpose

This deliverable analyzes one of the major conditions (extracting data from Endomondo) that could potentially lead to failure of the project.

Objective

This brief report identifies major steps involved in collecting and extracting user work out history from Endomondo site. Next, in the form of a table, the report describes all activities that are involved in extracting data specifically from Endomondo, the different approaches and solutions discovered so far as well as the steps necessary to control and manage this major risk.

Report

Major Data Extraction Tasks

1. Identifying source of data extraction.
2. Test and validate authentication scheme used by data source.
3. Test and identify HTTP request and response structure used by data source.
4. Identify format of Data Schema to be extracted.
5. Enlist networking and JSON parsing libraries

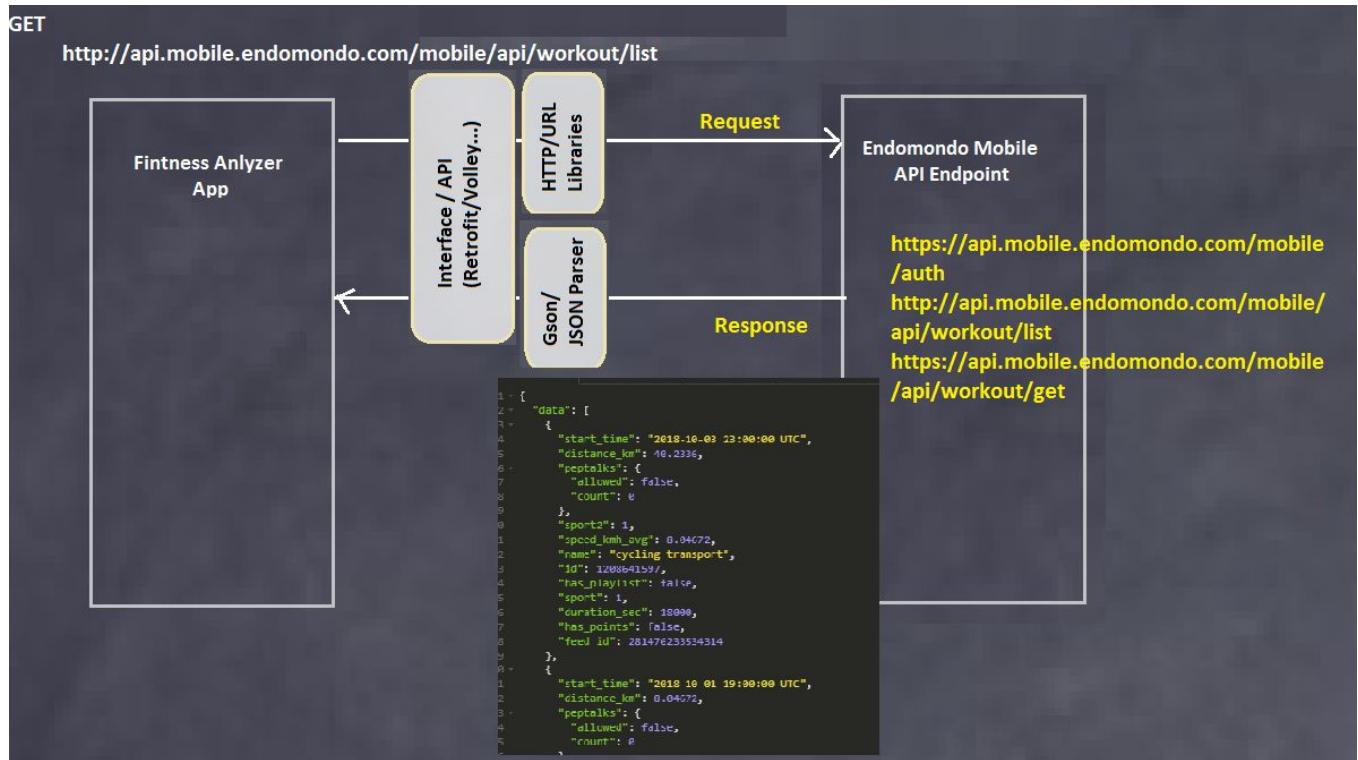
Findings, Task Status & Brief Explanation

The task of data extraction involves several steps. The table below identifies all the steps and the team's findings related to each step along with brief explanation and status of each task.

	Task	Findings/Brief Explanation	Status
1	Identify Information source.	<p><u>First Option (WEB REST API):</u></p> <p>https://www.endomondo.com/rest/session</p> <p>https://www.endomondo.com/rest/v1/users/UserSpecificId/workouts/history?limit=500&expand=workout</p> <p><u>Second Option (Mobile REST API):</u></p> <p>https://api.mobile.endomondo.com/mobile/auth</p> <p>http://api.mobile.endomondo.com/mobile/api/workout/list</p> <p>https://api.mobile.endomondo.com/mobile/api/workout/get</p> <p>Both API's are active, however no official documentation available for both API's.</p>	<input checked="" type="checkbox"/>
2	Identify Authentication Scheme used by REST Endpoint.	<p>Several authentication schemes available for restful web services such as: Basic Authentication, oAuth v1/v2, Digest Authentication, Token Authentication. Challenge was to identify the authentication scheme used by Endomondo.</p> <p>Testing with development tools like Insomnia and Postman, our team successfully authenticated using token authentication to Endomondo mobile API identified in task 1.</p>	<input checked="" type="checkbox"/>
3	Identify structure of HTTP requests and responses sent back & forth between two end points.	<p>Testing with developer tools (Insomnia & Postman), the structure of HTTP requests and response were identified.</p> <p>An HTTP Post request is used to authenticate by passing in valid Endomondo credentials as well as some other required parameters as HTTP headers. The response from Endomondo mobile REST API includes user Endomondo profile related information including the following key piece of information:</p> <p>Authentication Token (used in every subsequent HTTP request)</p> <p>The token received is added as a header in subsequent calls to Endomondo API to retrieve user workout history. The response from Endomondo, as shown in screenshot that follows is data in JSON format:</p>	<input checked="" type="checkbox"/>

		<p>200 OK TIME 343 ms SIZE 1472 B</p> <table border="1"> <thead> <tr> <th>Preview</th><th>Header 14</th><th>Cookie 2</th><th>Timeline</th></tr> </thead> <tbody> <tr><td>1 - {</td><td></td><td></td><td></td></tr> <tr><td>2 - "data": [</td><td></td><td></td><td></td></tr> <tr><td>3 - {</td><td></td><td></td><td></td></tr> <tr><td>4 - "start_time": "2018-10-03 23:00:00 UTC",</td><td></td><td></td><td></td></tr> <tr><td>5 - "distance_km": 48.2336,</td><td></td><td></td><td></td></tr> <tr><td>6 - "peptalks": {</td><td></td><td></td><td></td></tr> <tr><td>7 - "allowed": false,</td><td></td><td></td><td></td></tr> <tr><td>8 - "count": 0</td><td></td><td></td><td></td></tr> <tr><td>9 - },</td><td></td><td></td><td></td></tr> <tr><td>10 - "sport2": 1,</td><td></td><td></td><td></td></tr> <tr><td>11 - "speed_kmh_avg": 8.04672,</td><td></td><td></td><td></td></tr> <tr><td>12 - "name": "cycling transport",</td><td></td><td></td><td></td></tr> <tr><td>13 - "id": 1208641597,</td><td></td><td></td><td></td></tr> <tr><td>14 - "has_playlist": false,</td><td></td><td></td><td></td></tr> <tr><td>15 - "sport": 1,</td><td></td><td></td><td></td></tr> <tr><td>16 - "duration_sec": 18000,</td><td></td><td></td><td></td></tr> <tr><td>17 - "has_points": false,</td><td></td><td></td><td></td></tr> <tr><td>18 - "feed_id": 281476233534314</td><td></td><td></td><td></td></tr> <tr><td>19 - },</td><td></td><td></td><td></td></tr> <tr><td>--</td><td></td><td></td><td></td></tr> </tbody> </table>	Preview	Header 14	Cookie 2	Timeline	1 - {				2 - "data": [3 - {				4 - "start_time": "2018-10-03 23:00:00 UTC",				5 - "distance_km": 48.2336,				6 - "peptalks": {				7 - "allowed": false,				8 - "count": 0				9 - },				10 - "sport2": 1,				11 - "speed_kmh_avg": 8.04672,				12 - "name": "cycling transport",				13 - "id": 1208641597,				14 - "has_playlist": false,				15 - "sport": 1,				16 - "duration_sec": 18000,				17 - "has_points": false,				18 - "feed_id": 281476233534314				19 - },				--				
Preview	Header 14	Cookie 2	Timeline																																																																																				
1 - {																																																																																							
2 - "data": [
3 - {																																																																																							
4 - "start_time": "2018-10-03 23:00:00 UTC",																																																																																							
5 - "distance_km": 48.2336,																																																																																							
6 - "peptalks": {																																																																																							
7 - "allowed": false,																																																																																							
8 - "count": 0																																																																																							
9 - },																																																																																							
10 - "sport2": 1,																																																																																							
11 - "speed_kmh_avg": 8.04672,																																																																																							
12 - "name": "cycling transport",																																																																																							
13 - "id": 1208641597,																																																																																							
14 - "has_playlist": false,																																																																																							
15 - "sport": 1,																																																																																							
16 - "duration_sec": 18000,																																																																																							
17 - "has_points": false,																																																																																							
18 - "feed_id": 281476233534314																																																																																							
19 - },																																																																																							
--																																																																																							
4	Data Schema to be extracted	As illustrated by the screenshot in preceding step, user workout history will be received in JSON format.	<input checked="" type="checkbox"/>																																																																																				
5	Network Libraries for connecting to Endomondo Mobile API	<p>Android Build-in library:</p> <ul style="list-style-type: none"> - HttpURLConnection (connecting to the endpoint) - URL - Java I/O (input streams to read responses) <p>Third Party Libraries:</p> <ul style="list-style-type: none"> - OkHttp - Google's Volley - Square's Retrofit <p>JSON Parsers</p> <ul style="list-style-type: none"> - GSON - JSON Parser <p>Although libraries have been identified, team is yet to explore the pros and cons of each and finally select the most suitable library.</p>																																																																																					

Data Extraction Illustration Diagram



Work Log

Asif:

Majority of the data gathering and extraction research including various authentication schemes was done by Asif. He worked on the use case related to the data gathering/extraction and was able to identify, and test authentication scheme in use by Endomondo API as well as the necessary HTTP structure for communicating with Endomondo endpoint.

As for the other group members, they had our respective use cases to complete. But, they contributed to the data gathering by searching for links and datasets related to the fitness data. This was done via Google Docs and Slack.

Airi:

Data visualisation. Most of the visualisation such as plotting the graph and the related diagrams are created by her.

Tung:

Parsing JSON files to data tables.

Extracting and processing the data. Providing data input for data visualization.

Ramy:

Produced a use case, sequence, and system sequence diagrams for the login process and researched links related to fitness data.