

Project DMC at DoS Requirement Specification

In the area of IT development, a requirement specification" is a document which contains a detailed description of the requirements and expectations needed to develop a system or a software application. It is a crucial step in the development process as it defines what the system should do and how it should behave. And also, which non-functional demands it must live up to, e.g. in the field of information and data security

Date: 18th of Sept 2023	Created by: IT dept. At DoS with MS STEs Annie Stahel and Niels Jespersen	Project manager: Ahlam Al_Rousan
	Version: 0.5	Systems owner: At DoS: Ahlam Al_Rousan

Version log

Version		Date	Summary of changes	Track changes?
0.5		18-21 sep 2023	First version and draft created during the mission	

1. Introduction

This project concerns the design and specification of a DMC at DoS, Jordan. The whole environment should comprise both the Core DMC as well as the data areas for external users (researchers, decision makers).

.2. Specs and demands

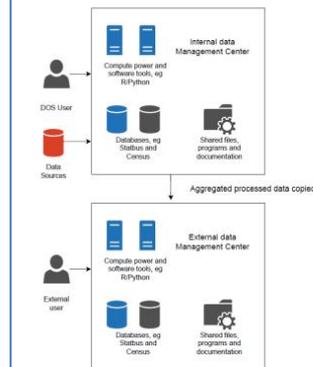
Describe the requirements for the system - as a minimum a list of requirements, but the requirements can be illustrated with use cases, user stories, screenshots, etc.

2.1 Functionel demands

What are the business defined goals and requirements that the system or application should fulfill – what business process is it supposed to support. What kind of users will use the system?

The system should be able to store and process data in a secure and centralized manner incl. the data that has been transferred from external sources and stakeholders. The whole statistical production process from data collection to data dissemination should be covered by the services of this data management center. The requirements were mentioned from early when the strategy was coined in February 2023 and comprise the following:

Data Management Center (DMC)



- The DMC centralizes data, statistical tools and computing power
- Security is heightened, because data is not directly accessible from desktop pcs.
- Data cannot be copied from the DMC to desktop pc's, ensuring that statistical production is not dependent on a single desktop
- Statisticians connect to a workplace inside the DMC using remote desktop
- Data Science work can take place alongside statistical production
- Tools standardized on SQL, R, Python
- Enough computing power (CPU, RAM) for all production processes.

2.1.A Data sources: data inflow and data “outflow”

All data that needs to be protected by DoS should be stored in the DMC.

Ingoing data can be:

1. Data from the enumerators / field workers
2. Data from MODEE, either as a transfer from other institutions or extracted from MODEE directly
3. Historical data does not represent data transfer but should be stored and kept inside the borders of the DMC

The data inputs to the DMC are sources coming via MODEE into DoS by IBM MQ service – or DoS can access data directly at MODEE. The data is in many cases loaded into the Oracle databases. These data streams can be preserved more or less the same way when the DMC is in place. Data transmission must always be supported by these procedures – USB sticks and email based data transfer is not allowed.

A gateway to the DMC for incoming data must be established.

Outgoing data transfer / dissemination:

The outgoing data flows are in most cases concerning data on an aggregated level, i.e. non-sensitive data. The data flows are e.g.:

1. Data needs to be published e.g. on the website <https://dosweb.dos.gov.jo/#> (The Web DMZ)
2. In a later step, data can also be sent to the Data Portal (not existing right now – no URL; but could be a part of the Databank environment in the future (The Web DMZ)).
3. Data is also transmitted to the Databank at <https://jorinfo.dos.gov.jo/Databank/pxweb/en/> (The Web DMZ)
4. Another publishing platform can be found (Arabic only) at https://jorinfo.dos.gov.jo/View_ind/

-
5. Data also needs to be prepared and transferred to the environment where researchers and decision makers can access data via specific interfaces. Researchers and decision makers should never be allowed direct access to the DMC Core but should have a separate environment; it could be in the form of a data warehouse or another kind of platform.
 6. External users will sometimes need data in the form of CSV or Excel files....these can be sent to the users by e.g. email. This data will always be aggregated data and not on a microlevel / personalized level. Request for data can be submitted via the website of DoS where the data request will be given to the responsible person / manager subject matter and then the data will be prepared and sent to the applicant via e.g. mail. This is in many ways a manual service process which need not be changed in the initial phase – but a more elegant solution would be to establish a platform where the requester can access a space and download the requested data.

Outgoing data on non-aggregated level / microlevel / personalized level will need to be treated in another way. This option will be dealt with later on (it will not be discussed during this mission in September 2023).

2.1.B Files and file formats / organization of files

2.1.C Users / different users and their needs for access to the DMC

Internal users: Internal users are the kind of users that have access to the DMC Core and can process data in one way or the other.

They must all be DoS employees in order to gain access to the DMC, and can be further subcategorized into segments:

1. Administrators in the IT department with Admin-privileges
2. The rest of the IT employees in IT department without Admin-privileges
3. The operators of the system themselves with the privileges needed for this. The IT Director will authorize the personnel in category 1, 2 and 3.
4. Users in the statistical production (the Statisticians). These do all have different privileges related to the subject area that they are working on and are experts in. The different data owners in the statistical divisions will authorize the statisticians with a need for access.

External users: You can decide whether users get a permanent access or a temporary access. External users will never be allowed into the DMC Core itself but will need to be serviced in a separate environment. An important question to decide is whether external users are allowed to take data "home" and download it to themselves or will the environment be closed for download and only visual inspection is allowed. If download of data is allowed, DoS should consider human resources needed for inspecting and enforcing the security procedures.

1. The field workers (enumerators) working with surveys and questionnaires in the field are not considered as internal users as such, since they deliver data to the DMC but they will not process the data. The survey process with providing data for enumerators and their questionnaires should be a separate process and described as a separate process. They will be authorized to read and write to only very specific tables in a database and the manager of the survey will be responsible for this authorization.

-
2. Any person or institution that needs access to insight in the data from DoS. Which does not entail that they get direct access to the data processing zone itself in the DMC, but rather that they get access to a separate environment where they can see and work with data either on an aggregated level or on a microdata level (personal data). These different users could be:
 - a. e.g. the Ministry of Labour and other decision makers will have a username and password and log in to see a part of data that is relevant for this ministry. Consider whether they will be satisfied with prepared files that are placed on a secure FTP-server. His Excellency DG will authorize external users in connection with a data agreement and security guidelines with the institutions.
 - b. Researchers and professors from e.g. universities. They are considered advanced users. They need access to data across different subject fields. Same setup as in a.
 - c. Administrative employees from ministries and agencies like e.g. MODEE are considered (ordinary) external users. Same as a.
 - d. MODEE is also a potential operator where data could be stored some time in the future. There is not enough knowledge at this point of time to specify the authorization process.
 - e. The citizens of the Jordanian population will not be authorized users – they can access the services from the website.

SIsI

2.2 Non-Functional demands

Be aware that there are many different types of requirements:

- *Functional requirements as above*

- *Data requirements*

What kind of data should the DMC be able to handle / and in what (file)format?

What kind of processes are applied to which kind of data?

What kind of data types does the DMC support / what does "data type" mean in this context (data type is e.g. integer, string of text)

What

- *Interface/communication requirements*

- *System performance requirements*

- *Quality requirements*

- *Design requirements*

- *Security requirements (with a specific section below)*

- *User interface requirements:*

Several user interfaces and technical set ups are needed

DoS needs a two factor logon system for remote access to the DMC when users are not present at DoS but will also make use of username and password:

1. The internal users can log on with username and password to the DMC when they are present at DoS premises.

2. The internal users will log on with two factor logon to the DMC when they are not present at DoS premises.

The services separated from the DMC:

3. The external users will always log on to environments with statistical data with two factor logon no matter where they are.

a

Number	Description	Requirement posed by	Priority
1	User tools / which types of tools should the DMC support? Se section in other document		
2			
3			
4			

2.3 Security requirements

This section should include the requirements for IT security, data security and data confidentiality. Consideration can be given to e.g.:

- Access management
- Roles and rights
- Logging
- Password management
- Handling sensitive information
- Sharing of data

Number	Description	Requirement posed by	Priority
1			
2			
3			
4			

3. Estimation (time and cost)

If possible, break the task down into parts that take about a week and estimate them. Suggestion: Use three-point estimation. Estimate in hours or weeks

The capacity estimation concerning the needs for servers and their configuration with regard to cores, RAM, storage and so and could be carried out with the help of a specialist consulting company with expertise in capacity analysis and capacity needs. The Danish experts recommend taking contact to a local company with expertise in capacity planning and configuration.

Mr. Niels has some recommendations in this area – but a more low level estimation cannot be carried out fully at the moment.

4. IT Architecture

Describe the overall IT architecture – if possible, provide a minimum a system diagram or a reference to it.

Software and tools needed for the Statisticians in DoS in the DMC:

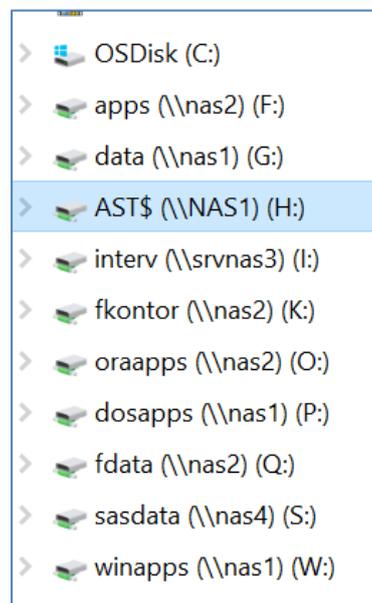
SPSS	Licenses need to be purchased + tested
Excel	No problems as far as we know – but perhaps Excel is not the best tool for statistical purpose
Stata	Check to see if it can run on the terminal server – NJN will help with this
R	Tool to come (what about competencies in DoS and IT department?)
Python	Tool to come (what about competencies in DoS and IT department?)
Toad, DataGrip, Sql-developer	These tools should also be placed inside the DMC and let Statisticians work with these database tools where the databases are.

DoS IT department has a vision of working with a standardized set of tools.

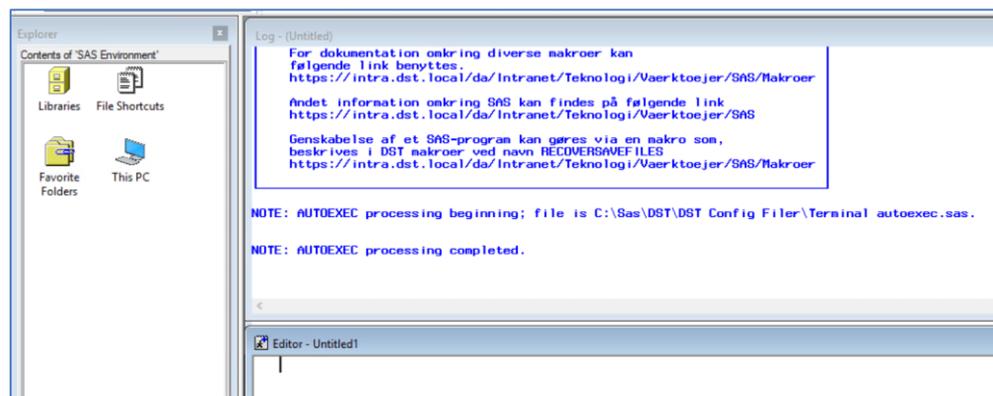
With regard to the Excel, the Statisticians can work with VDI connecting to the DMC and working with Excel and other tools that are not on the terminal server installation.

From a laptop, connect to the terminal server and work with the tools.

Statisticians should not have data resident on their laptops – statistical data on the C-drive of the laptop should never be allowed.



Start the tools on the terminal server:



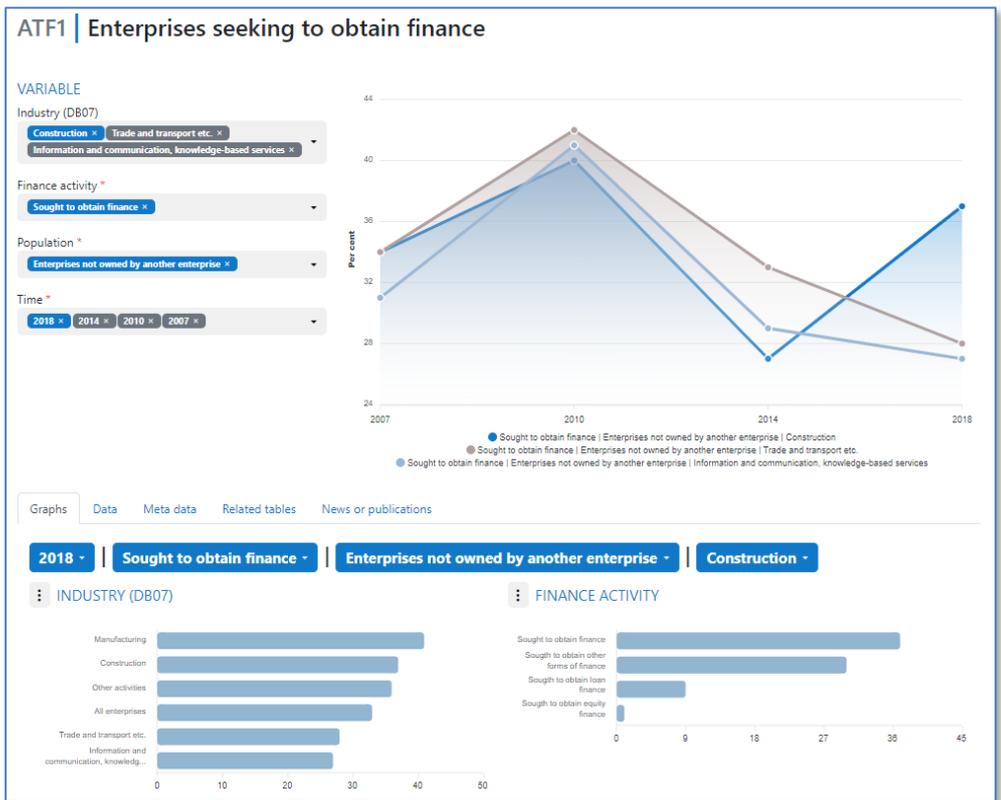
We recommend that DoS write down use cases. Ms. Charlotte Nielsen is recommended to host a workshop about the scenarios:

1. I am an employee at DoS. I want to log on and get access to my tools....[fill out]
2. I am a researcher at DoS – this is how I log on and these are my options.....[fill out]
3. I am a governmental official – these are my access rights and this is the environment that I log into.....[fill out]
4. I am a citizen from the public / Jordan (and these users may just get access to the public website)[fill out]

The needs, requirements and wishes of the researchers, governmental officials should be identified.

Could they e.g. use a presentation window and a tool that will give them dynamic and flexible view of a selection of the data from the data bases at DoS?

It could be based on Vue.js and could look something like this:



The World Bank suggests that the analysis section for the governmental officials and decision makers should also contain written analyses and narratives so that there are insight models for different expertise levels in the users.

5. Presentation and user interfaces

The login pages for the different users should be or could be set up in various manners. The employee at DoS will have to choose a different web page or ip address that will direct them to the resources that their privileges will grant them access to. Other users may have other options in the log on interface.

6. Program design

If there are changes to programs, they should be briefly described and reference is made to program documentation.

7. Database design

Describe the need for databases og any changes to existing databases and refer to documentation of this.

At the moment, DoS has 4 different Oracle databases and 2 PostgreSQL databases plus a number of filebased datasets.

8. User manual

A few words about the coming or existing user manual or guide. Please refer to the updated user guide, if it exists. Can be attached as an appendix.

9. Test

In short words, describe how the testing will take place. Which tests, which test data, which test environment, which test tools...including test cases etc.