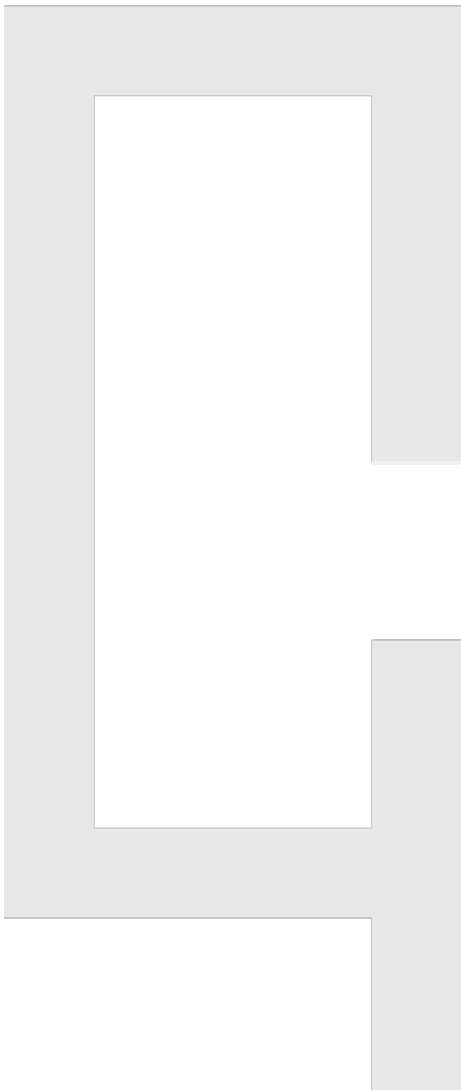


FINAL REPORT

DAta Maturity Model Utilization

V.2

2014-01-10



FINAL REPORT	1
SUMMARY	FEL! BOKMÄRKET ÄR INTE DEFINIERAT.FEL! BOKMÄRKET ÄR INTE DEFINIERAT.3
DATA MATURITY MODEL: BACKGROUND AND PURPOSE	3
DATA MATURITY MODEL: DESCRIPTION OF CONTENT	3
DATA MATURITY MODEL IN ITRACT	7
Identification of data sets to be used in ITRACT services	8
Identification of data sources and creation of a data set matrix per data set.	8
Defining the characteristics of a data source using the DAMM specification.	9
Appendix 1: Template for Activity 2 – Data Set Inventory	9
Appendix 2: Template for Activity 3 – Characteristics of the Data Source	9
THE DEVELOPMENT PROCESS OF DAMM	10
DAMM AND TRANSFERABILITY	10
EVALUATION AND DAMM REDESIGN	11
CONCLUDING REMARKS	11
APPENDIX	12

Introduction

The Data Maturity Model (DAMM) can be useful in several ways in service design development processes. This report will summarize the intentions behind the model, how the model can be used, and how it has been utilized in the ITRACT-project. In ITRACT, a Data Maturity Model is especially valuable since it addresses the *service transferability deliverable* by simplifying the service description of specific services regarding required and available data sources and utilization to implement digital services across regions. In particular, the Data Maturity Model addresses the question how different open data stakeholders can benefit from performing systematic open data assessment (Hjalmarsson, Johansson and Rudmark, 2014).

Data Maturity Model: BACKGROUND AND PURPOSE

The initial idea with DAMM (Data Maturity Model) in the ITRACT project, was to support actors, preferably system developers, in publishing data by using a specification to indicate the level of data maturity. Existing frameworks, e.g. *five stars of open data*, was seen as an insufficient support as indicator of data maturity service transferability and that was considered as one argument for why the development of DAMM was required in the project.

DAMM: two initial purposes

The *initial* purpose with DAMM was to enable transferability of services between regions by the ***identification of available data sources***. From a transferability perspective, data sources were seen as critical components due to both availability and transferability. Aspects such as differences in licence agreement, data format etc., are examples of aspects that possibly affects the design and transferability of digital services. One important aspect in service design and service realization is that relevant data sources are available and possible to utilize by service developers. Nevertheless, data sources are needed for the realization of digital services, why data maturity model (DAMM) was proposed as a tool to support the identification of availability of data sources and further data source utilization.

The *second purpose* was to support the ***identification of data sources needed to realize a specific service***. This is somewhat the opposite from the more general *identification of available data sources in a specific region*, as a specific service drives the identification of specific data sources. Hence, this analysis is also based on the similar specification structure and the same aspects are considered important. More about these two appendixes (i.e. specification support) are described in the following sections.

Data Maturity Model: DESCRIPTION OF CONTENT

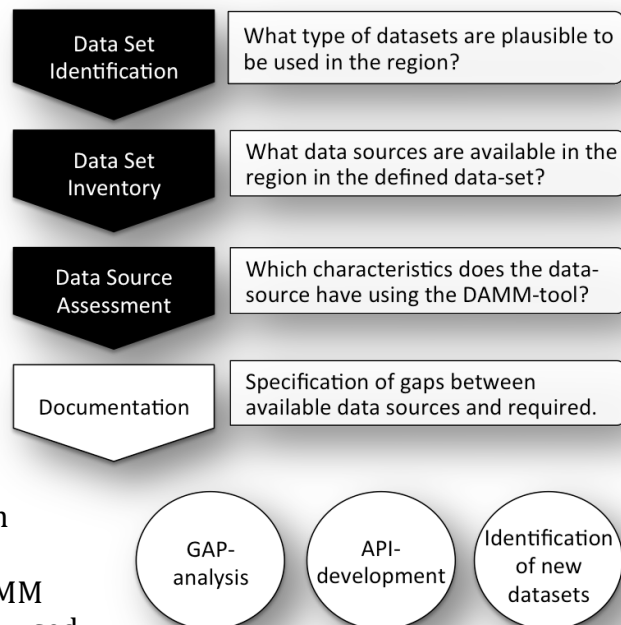
A Data Maturity Model is a way to simplify the representation of the “state of things” regarding availability of data (related to personal transports in the ITRACT project) in a specific region. Further, the identification of available data sources can be used to compare regions regarding how a selected service can be implemented, which actions that needs to be taken into account, in order to successfully implement selected services in a specific region.

The Data Maturity Model includes support for the following activities:

- Preliminary (and iterative) assessment of required data sources/data sets
- Inventory and classification of data sources/data sets from a categorization
- Description and classification of the identified data sources/data sets
- Documentation of data sources/data sets

The activities that DAMM can be used for, shown in figure 1: DAMM – Activities and Purposes below, are not necessarily sequential. Different activities are relevant for the purpose of determining the maturity of Data with regards to accessibility and transferability in a specific region and other activities are more suitable for data categorization and specification as such.

Although, the activities in DAMM are interconnected and could be used as a method to perform GAP-analysis and/or Requirement Elicitation for API-development and Data Set Identification.



The *Data Set Identification* can be used as a basis to identify available data sets. Further, *Data Set Inventory* can support the specification of available data sources for specific regions and defined data sets, which can be seen as a more detailed description of Data Sets. This analysis, using the DAMM specifications, can then be used, alone or in combination, to conduct a GAP analysis. The result from such a gap analysis can identify the need for the development of APIs and/or new data sets.

The development of DAMM has been inspired by, and intend to complement existing methods such as: *European PSI Scoreboard (PSI indicators)* and *“The five stars of open data”*. DAMM provides a generic description of datasets and detailed domain specific specifications. In ITRACT, a Data Maturity Model is especially interesting since it addresses the *service transferability deliverable* by simplifying the service description of specific services regarding required and available data sources and utilization to implement services across regions.

To be able to differentiate and compare the usability of different data sources, there is a need to define a data source from the “providing data sets” of the data-source, and also the functional mode of the data provided. A data set in the DAMM-model is a way of clustering data, sharing the same context of use, or area of creation. Data sources in the data sets does not necessarily have to origin from one and the same data provider, but probably there are at least one dominant actor providing data sources for a specific data set. Examples of data sets with relevance for ITRACT are Public transport, Road administrator, Public services etc. To support the utilization of DAMM in a service design process, both as a tool to assess data maturity and level of transferability and to support data and service documentation, two appendixes were developed:

Appendix 1: an inventory of Data Sets to enable an overview/assessment of data sources available and/or required to realize a specific digital service. Each Data Set is described in a more detailed specification and required data sources can be characterized in terms of *level of openness, processability, technical availability, licence/ availability/ cost, level of support* and *level of quality*. These parameters have been identified as relevant to determine Data Maturity and enable assessment of a certain realization potential.

To enable a comprehensive overview of available data sources in a region the DAMM model categorizes data sources in functional modes. In the model four functional modes have been defined (*see Matrix 1: Data Source Characterization – an example*); STATIC data, DYNAMIC data, STATISTICS data and SERVICES (which consist of data processed or aggregated to provide a new set of data)

Data set	Static	Dynamic	Statistics	Services
Publ transport	Bus stop, Bus line (geographical), time table	Bus position, delays, disturbances	Bus line departure density, Bus line historic delays	Travel planner
Road administrator	Road network	Floating car data, accidents,	Road usage, transport types,	Congestion detection, parking advice

Taxi/Car pooling	Taxi stops, car register	Positions	Route statistics	Ordering, pre-ride price est.
Public service	Location of schools, hospitals etc., school curriculum	Available healthcare appointments	Demographics	Doctor's appointment booking

Matrix1: Data Source Characterization – an example

Appendix 2 provides detailed description of available and/or required data sources, based on a number of relevant parameters to ensure utilization of data sources required for a specific digital service. In order to describe a certain data set to enable the assessment of the potential for service realization and thereby the level of transferability of services across regions, the following parameters have been identified as essential to use: *Support, Access, License, Cost, Technical* and *Quality*. The documentation is regarded as a valuable basis evaluating possible realization and development of services in other regions. Each parameter is specified by describing a number of underlying variables using “YES”/”NO”. One integrated activity in the development of DAMM was data sources classification.

Type	Available (yes/no)	Short description	Contact info/URL
Completely Free	Yes: The full and most recent version of the API/dataset that can be retrieved/used without any costs whatsoever.		Link to information about this grant.
Free with restrictions	Yes: data is available for free, but under some kind of restriction, e.g. Only for a limited (trial) period, older versions of the data, a non-full dataset.	Description of the restrictions imposed on free access.	Link to information about the restrictions.
Associated with cost	Yes: there are versions of the data that need to be purchased. This doesn't mean that there are only pay-versions.	Description of pricing model and prices (if this can be described briefly).	Link to pricing information and conditions relating to this.

As mentioned, service developers are the primary user group of DAMM, and the intentions is to support them in the process of (1) analysing availability and maturity of available data sources for a specific region and possible for a specific domain, as well as (2) identifying relevant data sources for a specific service, (3) as a basis to assess a possible service realization, and (4) to document data sources required for the realization of specific digital services.

Digital Services that are specified according to DAMM specifications, functions as an important declaration and provides the developer with a “quick overview” of data sources required and the intended use.

Data Maturity Model in ITRACT

DAMM in ITRACT had two initial functions: (1) as a support tool for developers in the service design process, and (2) as a support tool to enable service transferability across regions. In the ITRACT project, the Data Maturity Model has been utilized by different actors and in several ways throughout the service innovation process. For the application in ITRACT, DAMM consist of a two-level assessment of data sources, to be used in some of the participating regions (See Figure 2: DAMM in ITRACT) and is outlined as a three activity process.

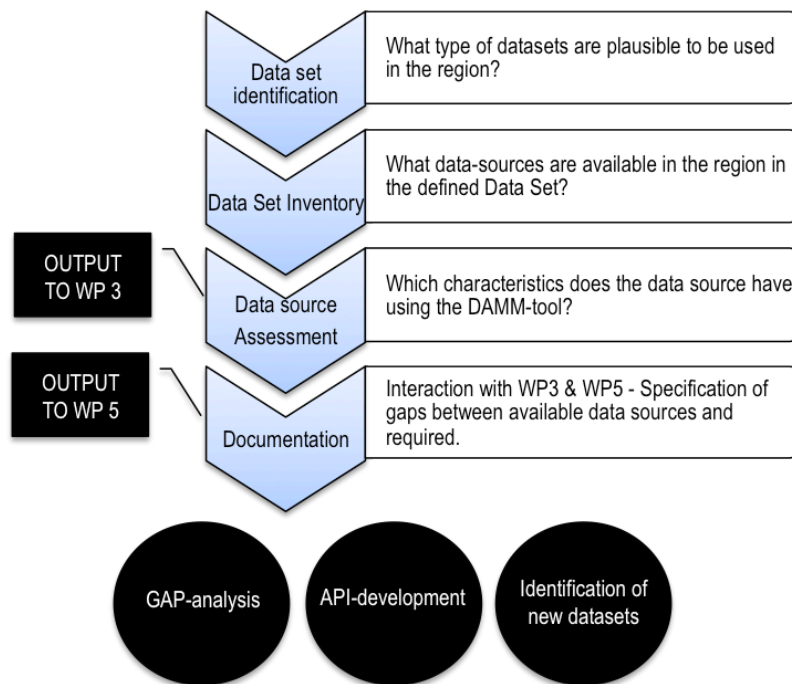


Figure 2: DAMM in ITRACT

In ITRACT, the intention was to use the Data Maturity Model in WP4 (Information Architecture and exchange mechanisms) to get a better understanding of available and needed data. The output should then be used for WP5 (Pilot actions on transport and accessibility) to assess which data that are available for the services that are about to be developed, and also by WP4 to be able to identify gaps between desired data available, and actions needed to be able to deliver the data required in WP5. The aim of the matrix is to give a comprehensive overview of relevant data sources available in a certain region.

The procedure starts by identifying which organizations or bodies that are responsible for the specific data set. There might be more than one organization providing relevant data in the specific context. Data sources available should then be categorized and the output should be described in “Data Set Inventory REGION” per data set (appendix 1).

Identification of data sets to be used in ITRACT services

The first activity aims to identify and categorize the different data sets that will be needed to identify and describe in LEVEL2 and LEVEL3. The aim with the *Data Set Identification* is to identify the needed data sets for a certain service in the region. In the first iteration each region was intended to include “Public Transport” and “Road Administrator” in the selected Data Sets. A region could, as a part of the service innovation workshops in WP3.1, identify the need for further data sets. The output is a description made using the template “Data set description DATASET_REGION.xls”.

Identification of data sources and creation of a data set matrix per data set

The second activity aims at identify and categorize data sources needed and to be able to create a comprehensive overview of available data sources in a region. As mentioned in the generic description of DAMM above, the model categorizes data sources in four functional modes (see figure 3: functional modes) and these were also the ones used in ITRACT to describe data sets;

- STATIC data
- DYNAMIC data
- STATISTIC data
- SERVICES, which consist of data processed or aggregated to provide a new set of data.

Data-set	Static	Dynamic	Statistics	Services
Publ transport	Bus stop, Bus line (geographical), time table	Bus position, delays, disturbances	Bus line departure density, Bus line historic delays	Travel planner
Road administrator	Road network	Floating car data, accidents,	Road usage, transport types,	Congestion-detection, parking advice
Taxi/Carpooling	Taxi stops, car register	Positions	Route statistics	Ordering, pre-ride price est.

Public service	Location of schools, hospitals etc., school curriculum	Available healthcare appointments	Demographics	Doctor's appointment booking
----------------	--	-----------------------------------	--------------	------------------------------

Figure 3: Functional modes in ITRACT

The aim with the matrix is to give a comprehensive overview of relevant data sources available in a certain region. The activity starts by identifying which organizations or bodies that are responsible for the specific data set. There might be more than one organization providing relevant data in the specific context. The available data sources are then categorized using the excel-specification: "Data Set Inventory REGION" per data set (Appendix 1).

Defining the characteristics of a data source using the DAMM specification.

The purpose with the third activity of the model is to describe each data source according to the predefined set of characteristics and with a common set of references. This gives the developer a possibility to quickly assess the maturity of a region for a certain service or application, and also the level of transferability regarding future and/or existing service to another region. It also enables an assessment over what measures that needs to be taken into account in the certain region for being able to deploy a certain service or application. The data source is described from six different perspectives; *Support, Access, License, Cost, Technical and Quality*. In line with the generic description, each perspective is described in a set of Yes/No questions, which comprises the most important aspects from a service developer's point of view.

The aim of the Data Source Characteristic Description is therefore to give a comprehensive overview of the data source according to the six perspectives. A guideline provides support for how each characteristic should be interpreted. The output from this activity is a characterization using the template "Data source characteristics DATA SOURCE REGION".

Appendix 1: Template for Activity 2 – Data Set Inventory

The purpose with the template is to provide a frame for documenting available data sources in an identified data set. This will give the project a comparability of data sources between regions. A template is included in this documentation named "Dataset inventory_DATASET_REGION.xls"

Appendix 2: Template for Activity 3 – Characteristics of the Data Source

The purpose with the template is to provide a common set of characteristics to describe available data sources in a region. The characterization is performed using a template document named: “DAMM-dataset-characteristics_DATASET_REGION.xls”.

THE DEVELOPMENT PROCESS OF DAMM

Viktoria Swedish ICT has been the leading actor in the development of DAMM and had two consultants as resources in the development team. Two workshops were held to elicit the basic requirements on DAMM. During the workshop relevant parameters for the description of data sources were identified, and grouped based on the two initial purposes of DAMM described in the first section of this report. The first part, described as Appendix 1, focused on parameters to support service based description of data sources for available data sources and/or for a specific digital service grouped in data sets. The second part, described as Appendix 2, focused on detailed information on the utilization of specific data sources.

Then DAMM specification was tested and evaluated during the development phase, through the utilization of DAMM as a tool to describe available and open regional transportation data sources from the city of Gothenburg, Västtrafik and Swedish Transport Administration. The evaluation phase resulted in some improvements and DAMM was redesigned according to these proposed changes.

DAMM AND TRANSFERABILITY

In order to enable transferability of services across regions, a standardized description of data sources is required. In the ITRACT project one such standard description has been developed, utilized and evaluated and named DAMM. Since the DAMM specification was not the primary deliverable in the project, it had to follow the overall innovation development process.

However, transferability is not only applicable to the data sources and data sets available, used and/or required, it is also about the innovation process as such. Therefore, we argue that a similar model should be developed in relation to the innovation process, to even further support the transferability of services across regions.

DAMM can be seen as an enabler for transferability by presenting a developer, or other actors, with an overview of relevant aspects to consider for a specific digital service. This overview will give the developer an opportunity to understand the terms of service realization and can possibly speed up the implementation process. Questions like - *is this service possible to implement in*

my region, what data sources are required, what other aspects are needed (terms of conditions etc.) to realize the service idea? - can be answered by utilizing DAMM. Besides standardized descriptions and categorizations, transparency in service documentation is an aspect that we believe enables transferability.

EVALUATION AND DAMM REDESIGN

Besides the integrated evaluation activity as part of the DAMM development phase, a separate analysis was made based on the application of DAMM in ITRACT.

As can be read in the section about the application of DAMM in ITRACT, all the regions did not apply the model for all the intended purposes. Due to the fact that the development of DAMM had to consider the overall service innovation process, the application of DAMM had to be adopted in the primary innovation process.

Service developers applied DAMM in some of the activities in the development process, such as a support tool to describe required data sources as a part of the documentation phase. The documentation can and have been used to support the documentation process as such as well as supporting service developers from other regions to assess the possibility to realize a similar service in their region. For this purpose, DAMM has been considered as a useful tool. One reflection made is whether the specification format could be improved/ changed to also suit non-technical actors in the development process. Another possible area of application for DAMM could be as a checklist for service developers, while assessing a specific data source. DAMM identifies important dimensions for a service developer to consider when evaluating availability and possible utilization of a specific data source and thereby the level of transferability of a service.

Another area of application identified, is for data owners to declare a specific data source. This can be useful both in a specific service development process, where the terms of utilization can be outlined, but also from the perspective of sustainability of data sources, where changes can be easily identified and described. This can be seen as one important aspect in the quality of data sources utilization.

Even if the application of DAMM in ITRACT has been limited, we believe that DAMM has been proven to be an efficient support tool for both the service innovation process and to enable transferability in digital services across regions.

CONCLUDING REMARKS

In the future, when DAMM is available from the start, it can be utilized to its full potential and thereby support the design service process to an even



greater extent. Important then is to integrate the application of DAMM in the overall service innovation design process.

APPENDIX

Found in separate folder.