SUPPORT INFRASTRUCTURE MODELS FOR RESEARCH DATA MANAGEMENT

DELIVERABLE: D4.1

Document identifier:	**evaluation_framework_v1.0.docx**
Date:	**25/01/2013**
Activity:	**T4.1**
Lead Partner:	**NIIF**
Document status:	**internal proofreading**
Work package:	**WP4**

Table of Contents

Glossary

Abbreviation	Explanation
EF	Evaluation Framework
QL	Qualitative
QN	Quantitative
RDM	Research Data Management
WP	Work Package
FB	Funding Bodies
RO	Research Organizations

1 Introduction

This paper is the initial proposal of the Evaluation Framework (EF) deliverable D4.1, and is meant to introduce the initial concepts and the vision of Research Data Management (RDM) oriented evaluation. The following conceptual statements are believed to lay down the key ideas of the proposed evaluation model:

- The targets of EF are the granted scientific projects analyzed from the RDM point of view. The metrics defined in this document are applied to the projects rather than complex programs built on multiple cooperating projects.
- It has been suggested that EF should become a hybrid model since it operates with qualitative and quantitative parameters too. On the one hand, the quantitative parameters will result in evaluation scores. On the other hand the outcome of the qualitative measures will be textual summaries. The nature of the different measures will be abbreviated with QL and QN representing qualitative and quantitative measures respectively.
- Any higher-level abstractions, such as evaluating the RDM compliance of national/regional projects are created as the numerical average of individual projects' evaluations in case of quantitative measures, and are created as description summaries for qualitative measures.
- EF should set up an evaluation system that enumerates the projects' RDM compliance, and scores them with a single number if possible. The meaning of this number, say either a percentage, or an integer between 1 and 100, is the strength of the RDM support. Low numbers indicate low level of RDM adherence, while high numbers indicate that RDM disciplines are strongly supported in the particular proposal.
- The evaluation criteria should result in a 'package of suggestions', which would reflect to the work of WP3. WP3 will create intervention action items, so the different evaluation parameters should be similar to those of used in the intervention model. This mutual relationship means that EF should focus on those measures that are sensitive enough to be intervened, and also EF should point out which measures are smart enough to yield large intervention impact. Parameters that do not endorse synergies among the different models built in different WPs should not be incorporated into the model.
- The findings are based on the WP2's evidence gathering, namely the data collected through the survey and in some cases, additional interviews. EF should conform to the structure worked out in WP2, meaning that the different stakeholders are classified as follows:
    - researchers,
    - funding bodies,
    - publishers, and
    - national bodies.
- EF is compiled to analyze:
    - the projects themselves, and also

  - their environment (national regulations, technical background, realistic possibilities),

  while avoiding being country-specific. However, in some cases certain exemplifications could not spare mentioning a couple of country-specific features. This kind of neutrality is a vital point of this framework since it was designed to be universal.

- To treat country-specific, or regional evaluation criteria, a two-layer plug-in model is introduced. While the higher-level criteria represent universal parameters, lower-level criteria represent localized specifics. The model is normalized, so the lack of localized parameters should not lead to any changes in the scaling.
- The evaluation criteria are independent of each other. It is highly unlikely that sequencing could be set up among them. However, the model should stay open for further probable alterations in the future.
- The evaluation process is able to analyze unique cases only. It was not designed to deal with a bulk of projects all at once. Should there be a country's evaluation, first the single cases have to be analyzed, and the final conclusions could be made at the end of this process.
- The different evaluation criteria are numbered to support easy reference. The numbering is unique and conforms to those of other deliverables.

In the following sections an evaluation framework model adhering to the basic principles shown above will be introduced.

2 Proposed Evaluation Framework model

2.1 Attributes of the model

The Evaluation Framework (EF) is a dynamic model, it analyzes the projects in three different phases during the life cycle of the projects. First, the project is evaluated during the call, then through the execution period and at last, but not least at the end of the project. Note that according to the nature of the execution of the scientific projects, the evaluation should occur multiple stages throughout the execution. Certain questions can only be validated at certain stages. The EF tends to be flexible enough to cope with this phenomenon.

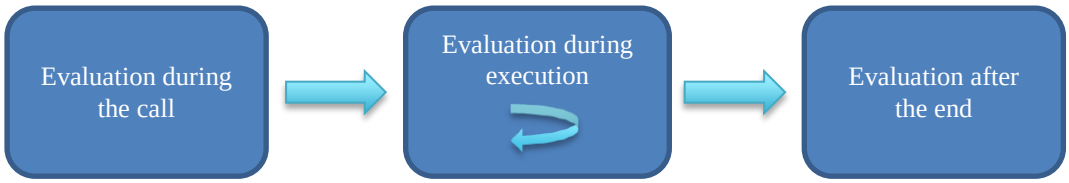

Figure 1: Evaluation framework is applied to projects in the different stages of their lifetime.

The Evaluation Framework is proposed to classify the evaluation parameters into the following parameter classes:

- Legal background;
- Organizational background;
- Technical background;
- User engagement.

Note that the evaluation model is applied to individual projects or programs run by research institutions, sponsored by funding bodies, controlled by national bodies, and exposed by publishers, being conform to the scientific ecosystem revealed in WP2.

The Evaluation Framework was designed to be measurable. This is an essential attribution, if we want to compare results in different environments. Nevertheless there are some measures that are either difficult or impossible to quantify, so certain criteria are qualitative, while certain criteria remain quantitative. So the following simple rules are used:

- The outcomes of QL measures are brief textual descriptions. These descriptions are treated individually, and are aggregated within, and for each parameter class.
- The outcomes of the QN measures are numbers determined by using the following rules:
    - The different QN measures are scored by integers from 0 to 100.
    - The whole proposal will also be scored by integers from 0 to 100.
- For each parameter a *preference number* is assigned. This number indicates the importance of the particular measure, and can start with 1, and can continue up to, say, 5. Low preference numbers indicate less relevant

evaluation parameters, while high numbers indicate evaluation parameters with higher relevance.

- Preference numbers are essential for weighting the individual evaluation parameters, and are also used for normalizing the aggregated measures. The normalization allows the model to take out evaluation parameters, and also to add new, e.g. regional, or country-specific evaluation parameters in the model without significantly abusing the proposed model.
- Weighted and normalized averages, as well as standard deviations will be used for describing the numerical evaluations. Weighted means represent the level of RDM support, while standard deviations indicate the level of balance among the different parameters. High standard deviation values point out the existence of small amount of dominant evaluation parameters.

2.2 Legal background

Scientific projects typically operate within a well-controlled legal environment that sets up the possible operational frames for them. This environment consists of different layers of law, from low-level governmental/ministerial regulations to high-level constitutions, or international/European law. The rationale behind inspecting the legal background is measuring the support level of RDM: it should not be too restrictive to forbid RDM-related activities; rather it should inspire project owners to increase their RDM efforts.

The proposed measures are as follows:

	Description	Type	Scores	Preference number
L1	Legislative support of RDM in the calls for proposal. Are there any legal obstacles to overcome when entering a call?	QN	0= no RDM-related regulations in the call. 100= targeted RDM-specific support in the call.	5
L2	Copyright regulations. Who does the Copyright Law consider as the author of such research data? Are there clear regulations to dissolve jurisdictional conflicts?	QN	0= no regulations to dissolve jurisdictional conflicts. 100= structured regulations to dissolve jurisdictional conflicts. Existing code of conduct.	5
L3	Regulation of the Data Protection Act (DPA) regarding the type of data and the transfer of research data. Are there special regulations, or separate chapters dedicated to RDM? How many?	QL+ QN	0= no RDM-specific regulations in the DPA. 100= targeted RDM-specific chapters, provisions in the DPA.	5
L4	Which law determines RDM-related matters? Are there any dedicated laws regulating this field? How it regulates?	QL+ QN	0= no RDM specific laws. 100= structured RDNM specific laws and policies.	5

Note that regarding RDM there could be difficulties (conflict situations, jurisdiction matters) in identifying the author, which could lead to problematic issues regarding copyright matters.

2.3 Organizational background

The organizational background measures attempts to determine the RDM-awareness, as well as the cultural impregnation of RDM within the organizations involved in scientific projects. There are two types of organizations inspected: on the one hand the funding bodies (*fb)* who initiate calls and create programs, and research organizations (*ro*) who run and realize the projects, on the other.

	Org.	Description	Type	Score	Preference number
O1	*fb*	Re-use of infrastructure in new projects, avoiding parallel operations.	QN	0= proof of parallel operations. 100= agreements regarding the re-use.	5
O2	*fb*	Number of funding bodies specifically supporting RDM.	QN	0= none. 100= a reasonable number of funding bodies support RDM.	5
O3	*ro*	Amount of new research grant income.	QN	0= there is no growth (compared to the amount before the proposal). 100= there is a significant growth (compared to the amount before the proposal).	5
O4	*ro*	Number of research dataset publications generated.	QN	0= none. 100= there is a significant growth (compared to the amount before the recent funding).	5
O5	*ro*	Number of research papers.	QN	0= none. 100= numerous research papers were published.	5
O6	*ro*	Number of RDM related organizations wishing to back up the technical introduction in during the incubation phase, for example: Grid, NREN.	QN	0= none. 100= numerous organizations support RDM.	5
O7	*ro*	Internal support: is there a policy within the organization compulsory to all which support research data management?	QL	___	5
O8	*ro*	External support: are	QL	___	5

		there any already existing initiatives, which handle the research data management between separate organizations?			
O9	*fb*	Sustainability approach. Are there any policies regarding this matter?	QL	0= no sustainability plans. 100= established sustainability plans.	5

Note that the National Research and Education Networks could fulfill this task since they already have the necessary infrastructure to implement this feature. Nevertheless, if there is a need for it, a brand new initiative could be set into motion.

2.4 Technical background

Inspection of the technical background focuses around measuring the infrastructure as well as the technical possibilities of project owners: software and other facilities that already support or would be able to support scientific data management. The measures try to reveal any possible bottlenecks that might handicap such activities. They technical background evaluation parameters are as follows:

	Description	Type	Score	Preference number
T1	Number of tools/equipment supporting RDM	QN	0= no software or hardware support. 100= two or three hardware and/or software tools used in RDM.	5
T2	Bandwidth	QN	0= large amount of sites with low bandwidth connection. 100= high bandwidth connections dominating.	5
T3	Storage space	QN	0= Gbytes of storage space. 100= at least Tbytes of storage space available for RDM purposes.	5
T4	Software support	QN	0= no RDM support software is used. 100= at least three pieces of RDM support software (e.g. file sharing) is used.	5
T5	Middleware support	QN	0= no middleware is used for RDM. 100= at least a middleware or some SaaS cloud is used for RDM.	5
T6	Percentage of potential user communities that take data deposit repository	QN	0= no such user communities. 100= 100% of the user communities.	5
T7	Number of downloads	QN	0= no downloads. 100= intensive scientific data downloads.	5

T8	Number of implemented cases	QN	0= no implemented cases. 100= two-three implemented cases.	5
T9	Quality of infrastructural state-of-art. Are there any specifications in the call that require high-level software and hardware support?	QL	___	5
T10	Key points: data transfer, data storage, accessibility of data. Do these key points play a major role in the proposal calls?	QL	____	5

2.5 User engagement

It seems that there is only a small scope of researchers, or groups of researchers who are aware of RDM issues, but these users are highly motivated in receiving more and more information about it. This set of measures tries to identify the user awareness of RDM, and also the number of people who already perform such activities. The users are divided into two groups: researchers (*r)*, who perform research, and staff (*s*) who are involved in executing and running for the scientific projects. The user engagement parameters are as follows:

	Users	Description	Type	Score	Preference number
U1	*r*	Number of users aware of RDM	QN	0= no interest. 100= active, motivated users, participation of related workshops.	5
U2	*r*	Increase in grant income/success rates.	QN	0= no successful proposals. 100= numerous winnings.	5
U3	*r*	Percentage of improvement in routine backup of data	QN	0= no improvement. 100= significant improvement.	5
U4	*s*	Percentage of improvement in range/effectiveness of research tool/software	QN	0= no improvement. 100= significant, at least 10% of improvement.	5
U5	*s*	Average time saved in RDM and grant proposal activities	QN	0= no time saved. 100= significant, at least 10% of time reduced.	5
U6	*s*	Average reduction in waiting time (time latency) for data requests	QN	0= no reduction. 100= significant, at least 10% of reduction.	5
U7	*s*	Users' brainstorming. Those who are aware of RDM, have they thought about a solution?	QL	___	5

U8	*s*	Do they have any suggestions or comments? Increased visibility of research through data citation.	QL	___	5

Note: The term ’user’ here does not mean the single researcher; moreover it means a group of researchers, virtual organizations like Grid virtual research communities.

3 Calculating method of quantitative parameters

The mean and the standard deviation values are calculated individually for each evaluation class. *LV-LD*, *OV-OD*, *TV-TD*, and *UV-UD* denote the evaluation values, as well as the standard deviations for legal, organizational, technical, and user engagement parameters respectively.

$$LV = \frac{\sum_{i=1}^{N_L} p_i L_i}{\sum_{i=1}^{N_L} p_i} \quad \text{(F1)}$$

$$LD = \frac{\sqrt{\sum_{i=1}^{N_L} p_i \left(L_i - LV\right)^2}}{\sum_{i=1}^{N_L} p_i} \quad \text{(F2)}$$

Here p_i indicates the preference values of the different evaluation parameters, while L_i denotes the individual evaluation values. *NL* denotes the number of parameters.

Similarly for organizational parameters:

$$OV = \frac{\sum_{i=1}^{N_O} q_i O_i}{\sum_{i=1}^{N_O} q_i} \quad \text{(F3)}$$

$$OD = \frac{\sqrt{\sum_{i=1}^{N_O} q_i \left(O_i - OV\right)^2}}{\sum_{i=1}^{N_O} q_i} \quad \text{(F4)}$$

In formulas (F3) and (F4) q_i indicates the preference values of the different evaluation parameters, while O_i denotes the individual evaluation values. *NO* denotes the number of parameters.

Similarly for technical parameters:

$$TV = \frac{\sum_{i=1}^{N_T} r_i T_i}{\sum_{i=1}^{N_T} r_i} \quad \text{(F5)}$$

$$TD = \frac{\sqrt{\sum_{i=1}^{N_T} r_i \left(T_i - TV\right)^2}}{\sum_{i=1}^{N_T} r_i} \qquad \text{(F6)}$$

In formulas (F5) and (F6) r*i* indicates the preference values of the different evaluation parameters, while T*i* denotes the individual evaluation values. *NT* denotes the number of parameters.

Similarly for user engagement parameters:

$$UV = \frac{\sum_{i=1}^{N_U} s_i U_i}{\sum_{i=1}^{N_U} s_i} \qquad \text{(F7)}$$

$$UD = \frac{\sqrt{\sum_{i=1}^{N_U} s_i \left(U_i - UV\right)^2}}{\sum_{i=1}^{N_U} s_i} \qquad \text{(F8)}$$

In formulas (F5) and (F6) r*i* indicates the preference values of the different evaluation parameters, while T*i* denotes the individual evaluation values. NT denotes the number of parameters.

Formula (F9) gives the sum of evaluation values for all parameter classes:

$$V = \frac{LV + OV + TV + UV}{4} \qquad \text{(F9)}$$

4 Conclusions

It is believed that this Evaluation Framework will highlight the weak points of the project calls and clearly mark those parameters where further intervention will be inevitable, based on the qualitative, as well as the quantitative measures introduced. To create intervention actions, evaluation procedures first the different funding schemes have to be understood.