
Prioritization and Action Items

Column Q has a (Y)es, if the parameter is quantitative. Column W has a (Y)es if the parameter can be weighted.

Parameter	Q	Rationale	Measurement	W	Comments
No. sources	Y	The relevance of entry <i>A</i> directly relates to the no. of entries depending on <i>A</i> .	Counting the links with <i>A</i> as source.	n/a	Can be determined based on SEE-IN LB contents. Is being used indirectly.
Strength of links	N	The relevance of entry <i>A</i> can depend on the strength of the links with <i>A</i> as the source. The strenght can be converted into a weight.	Link attribute.	n/a	Can be determined based on link attribute. Needs to be converted into a quantitative weight.
Relevance of targets	Y	The relevance of entry <i>A</i> can depend on the relevance of the targets of the links with <i>A</i> as the source.	Relevance of all target entries.	Y	Can be determined starting with those entries that do not appear as a source, and then working backwards.

number of requirements a dataset meets. For an entry A , we can define a local relevance r^{loc} as

$$r^{\text{loc}}(A) = \sum_{i=1}^{L_A} w_i, \quad (2)$$

where L_A is the number of links in which A appears as the source, and $w_i, i = 1, \dots, L_A$ are the weights assigned to each link. In the most simple case, we can set all $w_i = 1$. Taking into account the strength of a link (weak, strong, crucial), we could assign the values 1, 2, or 3 for weak, strong and crucial links, respectively.

The next level of complexity would take into account the relevance of the target. In this case, all relevances would become interdependent and would have to be determined in a recursive way. In this case, publishing a new entry and linking it to one other entry has the potential to change the relevance of a large number of other entries. However, at this level, the relevance would not just be derived from the immediate environment of an entry but reflect its global relevance. A measure for the global relevance r^{glob} is defined here as

$$r^{\text{glob}}(A) = \sum_{i=1}^{L_A} r_i^{\text{glob}} \quad (3)$$

Combining equations (2) and (3), we can define a weighted global relevance by

$$\hat{r}^{\text{glob}}(A) = \sum_{i=1}^{L_A} w_i \cdot \hat{r}_i^{\text{glob}} \quad (4)$$

Based on the global relevance of requirements, we can define the relevance $r(B)$ of a dataset or product B . If B meets the requirements $R_i, i = 1, \dots, K_B$, then the relevance of B is defined as

$$r^{\text{glob}}(B) = \sum_{i=1}^{K_B} \hat{r}_i^{\text{glob}}. \quad (5)$$

For SEE-IN KB entries:

$$\hat{r}^{\text{glob}}(A) = \sum_{i=1}^{L_A} w_i \cdot \hat{r}_i^{\text{glob}}, \quad (6)$$

L_A : number of entries E_i that are targets in the links with entry A being the source;

w_i : weight of the link between A and E_i ;

r_i^{glob} : global relevance of E_i .

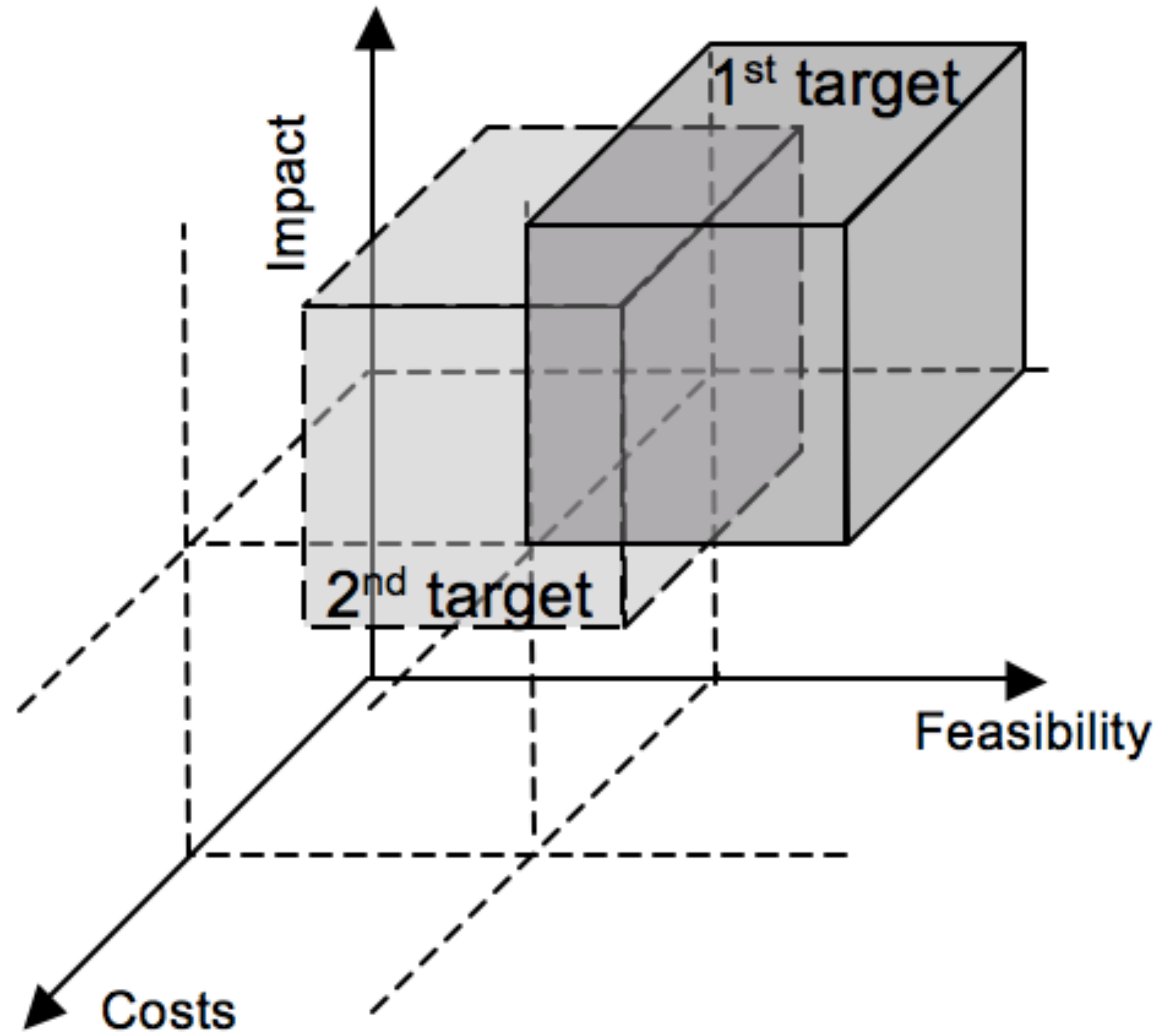
For external datasets or products:

$$r^{\text{glob}}(B) = \sum_{i=1}^{K_B} \hat{r}_i^{\text{glob}}, \quad (7)$$

K_B : number of requirement entries R_i that are met by dataset or product B ;

r_i^{glob} : global relevance of requirements R_i .

SEE-IN KB Prioritization



7.1 Specific Actions To Finalize Gap Analysis and Prioritization

In this section we list a number of actions that need to be taken in order to complete the gap analysis and make progress towards the goal of a prioritized list of gaps.

For TDT1, the following actions have been identified so far:

- Complete the identification of ESDGVs using the goal-based approach for the existing indicators and the identification of domain-specific EVs that match these ESDGVs.
- Complete the review of the targets with respect to additional indicators that would ensure that progress towards the targets accounts for environmental indicators.

For the completion of the documentation of the gaps, the following actions are pending:

- Merge the list of domain-specific EVs and the ESDGVs into one comprehensive and consistent master list of EVs.

- Ensure that all gaps are linked to proper EVs included in the master list of EVs. This requires a cleaning of the CGT and of all relevant links in the SEE-IN KB.

For the prioritization, the action required include;

- The algorithms for the ranking based on links has to be finalized and implemented in the SEE-IN KB.
- A more complete matrix of potential parameters to be used to define a set of metrics need to be compiled. Table 27 is a starting point. For datasets and products, continuity (length of time series) needs to be accounted for. The uniqueness of a dataset also increases the relevance. For a requirement, the frequency of its appearance in various databases indicates relevance. For both, datasets and requirements, user rating can be included as a component of relevance.

