Cost-Benefit analysis of becoming FAIR

Guidebook

DG RTD
Directorate-General for Research and Innovation

May 2018

pwc

DISCLAIMER

The views expressed in this document are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission.

The European Commission does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof.

The document contains a brief overview on how the cost-benefit analysis (CBA) mechanism should be used, for what it was designed, how to make use of it, how does it function, what are its limitations.

CBA FAIR Mechanism guidebook

Prepared by PwC EU Services for the "Cost-Benefit analysis of FAIR research data" project.

Version	Date	Modified by	Short description of changes
0.02	03/05/2018	PwC	Revision
0.01	03/05/2018	PwC	Creation of the guidebook

1 Introduction

FAIR Research Data CBA Mechanism

The FAIR principles are seen as potentially efficient but have not been fully implemented. For such cases, cost-benefit analysis are usually used to find efficient solutions to address issues that the market cannot solve itself.

With that in mind, the **Cost-Benefit Analysis (CBA) mechanism** primary objective is to identify scenarios where the implementation of the FAIR principles would result in greater benefits than costs. Indeed, according to the theory related to cost-benefit analysis, a programme should normally be adopted if the benefits are greater than the costs.

Therefore, this cost-benefit analysis will ultimately measure the efficiency of implementing the FAIR principles for a specific scenario by calculating and monetizing the costs and benefits incurred by an organisation, a business unit, a project, etc.

Besides, this cost-benefit analysis is part of a broader study conducted to estimate the impact of becoming FAIR – both in terms of costs and benefits.

This very document is intended to describe the **Cost-Benefit Analysis (CBA) mechanism** and provide guidance in order to make the most of it.

This document defines:

- The objectives and methodology of the CBA FAIR mechanism;
- The scope of the questionnaire;
- The instructions to be followed;
- Whom to contact in case of misunderstanding and/or misinterpreting.

FAIR Research Data CBA Mechanism

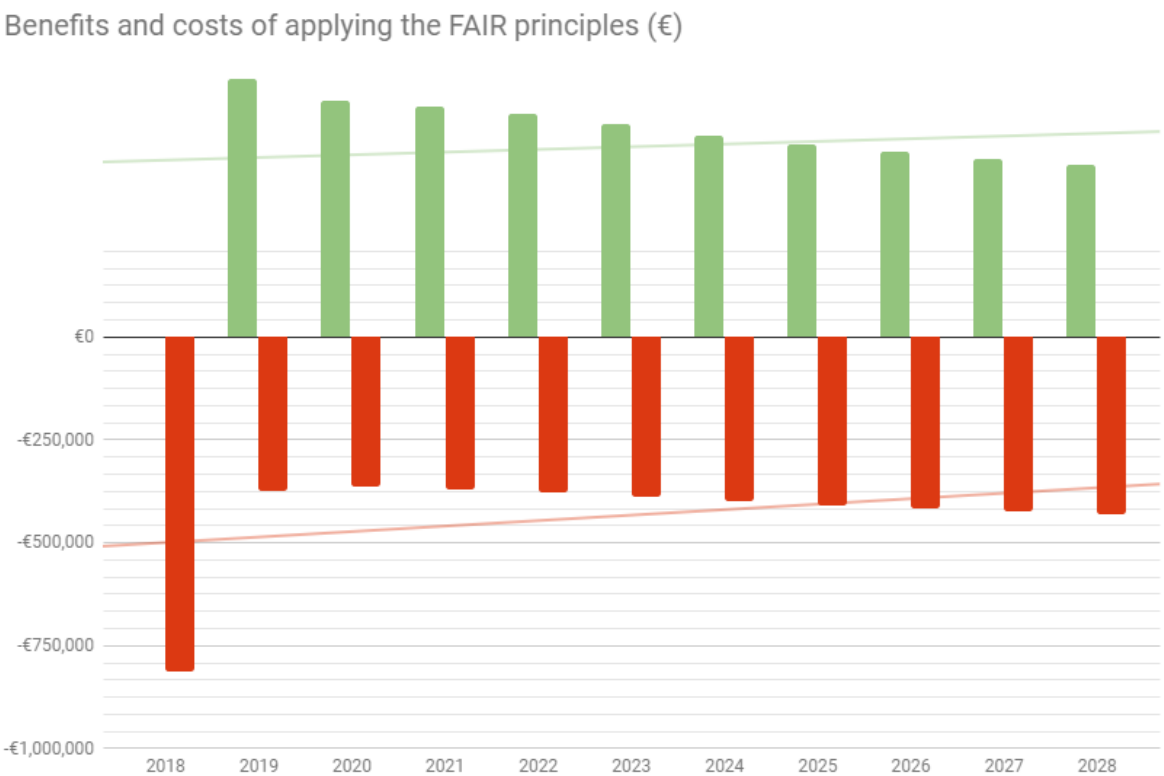
Why

The European Commission is conducting a study to determine the value of FAIR data, within and across scientific disciplines, both in economic and non-economic terms, and to contrast it against the current situation where a majority of research data is not adhering to the FAIR principles (i.e. un-FAIR).

From this study there is three underlying objectives:

- Obj01: To determine the cost of not having FAIR research data for the EU science and innovation system and as a result to the EU data economy
- **Obj02: To estimate the costs and benefits of FAIR research data, both in economic and non-economic terms**
- Obj03: To provide recommendations on the next steps concerning making research data FAIR

This mechanism adheres to ***Obj02*** and aims to depict cost and benefit of having FAIR research data.



FAIR Research Data CBA Mechanism

How

The **Cost-Benefit Analysis (CBA)** is one of the standard evaluation tools applied in the definition of the social and economic impacts in order to determine, in economic terms, whether a programme – *in our case the FAIR principles* – is worth implementing.
To arrive at a cost (*disadvantage*) **or** benefit (*advantage*) of fully adhering to the FAIR principles you will be asked to provide three types of input:

1. Inputs about your organisation, business unit, project, etc. ;
2. Inputs about your current compliance with the FAIR principles;
3. Inputs about the way you are currently conducting research, manipulating data but also the costs of storage, licencing currently faced.

As a consequence, the mechanism will interpret these inputs, identify them as costs or benefits and quantify them. Eventually, the costs and benefits will be monetized. In this last step, the net present value (NVP) will be calculated.

Who

The intended audience of the mechanism is mainly research performing organisations and research / data infrastructures as it is perfectly fitted for them. Nevertheless, other type of organisations may find it worthwhile. However, as the questionnaire is relatively oriented towards research data, the results coming out from the mechanism might be to some extent biased.

This mechanism is expected to be filled in by a person or a group of people in the organisation with the following characteristics:

SALARY CLEARANCE

Access to information regarding the salaries, the time spent on data related activities or the costs of data management practices (e.g. storage costs, licences fees)

DISCIPLINE UNDERSTANDING

Understanding of the specificities of the different research disciplines

Typically, research project managers or data management officers within research organisations, research / data infrastructures would be able to assess the costs and benefits of applying the FAIR principles. In any case, we advise you to have one person coordinating the assessment and gathering the right information.

Privacy matters

For the sake of accuracy of the model, actual value should be provided. In that case, all information provided in the questionnaire is considered sensitive and will not be publicly disclosed or used.

2 Instructions

FAIR Research Data CBA Mechanism

Case study workflow

The completion sequence of the mechanism will be as follow;

1. A version of the CBA mechanism will be shared beforehand together with instructions;
2. *(**Optional**) A first meeting will be held during which the mechanism will be described and preliminary questions will be answered;*
3. Filling in of the CBA mechanism excel sheet, according to the methodology described below;
4. *(**Optional**) If required a second meeting would be organised to answer any remaining question or provide any further assistance*;
5. Wrap-up meeting during which the results are presented and interpreted.

Fill in the questionnaire

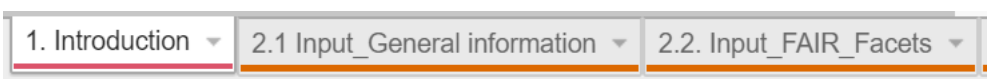
The CBA Mechanism is embodied in an Excel sheet. In this excel sheet you will come across four types of sheet as depicted below.

Part of the mechanism	Name	Tab content
1. Introduction	1. Introduction	General overview of the Mechanism
2. Input Data	2.1 General information	General information about your organisation or your project *(to be filled in)*
	2.2 FAIR Facets	General information about your compliance to the FAIR facets *(to be filled in)*
	2.3 Time	Information on how researchers spend their time *(to be filled in)*
	2.4 Miscellaneous	Information about storage, subscription fees, etc. *(to be filled in)*
3. Results	3. Results charts	Chart displaying the results of the cost-benefit analysis
	3.1 Results	Results of the cost-benefit analysis
4. Annexes	Calc_time	Technical sheet computing the value of time spent/saved.
	Calc_Misc	Technical sheet computing the value about storage, subcription fees, etc.
	Calc_Storage	Technical sheet computing the value about storage
	Glossary	Glossary defining the terms used throughout the sheet
	Researchers_Salaries	Technical sheet computing the salaries of researchers/per country

The four types are the following:

1. Introduction sheet, which gives an overview of the context and the mechanism;
2. Input sheets, which request the respondent to provide data;
3. Result sheets, which display the added value of FAIR research data based on the data previously entered;
4. Annex sheets, which provide further explanation to the respondent. Calculations and assumptions are also nested in the Annex sheets.

You can easily navigate through the different sheets using the tabs at the bottom of the excel file. Be aware that the annexes compile the calculations but does not requires any contribution from your end. Their role is merely informative.



FAIR Research Data CBA Mechanism

In the ***Input sheets*** you will be request to provide input. There are 4 ***Input sheets*** to answer to while only three types of inputs, as already mentioned:

1. Inputs about your organisation/project/business unit;
2. Inputs about your current compliance with the FAIR principles;
3. Inputs about the way you are currently conducting research and manipulating data.

As depicted below, you will be asked to fill a set of boxes in the ***Input sheets.***
The boxes to populate are already prefilled* with blue highlighted values and their background is set to light grey. The type of value to fill in is indicated at the left side of the cell.

#	2
#	4
EUR (€)	€12,000.00
hours	10
Profile	Research fellow

In the unfortunate case where you are not able to provide an answer, leave the cell as it is. The value already filled in will be used for the calculations.

Should you not understand a term or a question, you can refer to the glossary which primary use is to define terms and concepts employed throughout the sheet. For the sake of clarity, terms and concepts are sorted by sheet.

GLOSSARY

SHEET : 2.2. Input_Time

Creation and collection of the data

Creation and collection of the data is the step in the research life cycle where data is created (i.e. through observations, experiments or simulations), and potentially useful and existing data is found and obtained. This step also includes the performance of veracity checks on the data

Pre-processing and data cleansing

Pre-processing and describing the data is the step in the research life cycle where inspection and checks are performed to improve and ensure data quality. Data modelling and data cleansing also take place in this step.

Integration of the data

Integration of the data is the activity in the research life cycle where data from disparate sources is aggregated to form one homogeneous dataset that can be analysed. As research data volume is surging , data integration becomes necessary to be able to aggregate data from multiple sources (and from different formats).

Analysis of the data

Analysis of the data is the activity in the research life cycle where the data is processed with the ultimate purpose to extract useful information, to elicit insights and eventually to formulate observations and conclusions. Analysis of the data or data modelling is often performed using specific software's, tools or methodologies.

**The questionnaire will be already prefilled with value originating from previous studies. Feel free to update with your own values or leave them as they are.*

FAIR Research Data CBA Mechanism

Results

After the data is provided, the CBA mechanism excel sheet showcases the results in two different sheets. The first sheet is displaying two graphs:

- The first graph is computing the benefit against the costs of implementing the FAIR principles for your organisation, business unit, project, etc. over time.
- The second graph is computing the cumulative net present value over time.

The second sheet provides a detailed overview of the costs and benefits stemming from a likely implementation of the FAIR principles.

From the second sheet, the box depicted below should draw your attention. The box summarises the actual value of having the FAIR principles for your organisation, business unit, project. This is the end result of what would entail having the FAIR principles.

In other words, the net present value is today's value of the predicted Benefit/Cost of having FAIR research data.

In our case: 150,035.52 €

Whereas the Internal Rate of Return measure the estimated percentage return of implementing the FAIR research data.

In our case: 11.3 %

3 Contact details

FAIR Research Data CBA Mechanism

- Who to contact in case of trouble
  - Should you have difficulties to fill-in the questionnaire and/or difficulties to interpret the results please reach out to:
    - Athanasios Karalopoulos, Policy Officer at DG RTD – athanasios.karalopoulos@ec.europa.eu
    - Nicolas Loozen, Manager at PwC EU Services – nicolas.loosen@pwc.com