

Data Management Plan

WORKING DOCUMENT

Project Number: 732049 – Up2U

Project Acronym: Up2U

Project title: Up to University

HISTORY OF CHANGES		
Version	Publication date	Change
1.0	13.10.2016	<ul style="list-style-type: none">Initial version

1. Data Summary

The key objective of the Up2U project is to bridge the gap between secondary schools and higher education and research by better integrating formal and informal learning scenarios and adapting both the technology and the methodology that students will most likely be facing in universities. The project is focusing on the context of secondary schools. The learning context from the perspective of the students is the intersection of formal and informal spaces, a dynamic hybrid learning environment where synchronous activities meet in both virtual and real dimensions. Up2U is developing an innovative ecosystem that facilitates open, more effective and efficient co-design, co-creation, and use of digital content, tools and services adapted for personalised learning and teaching of high school students preparing for university. The project addresses project-based learning and peer-to-peer learning scenarios.

Up2U provides as part of the aforementioned ecosystem a learning management system (LMS) that integrates tools and applications provided by the project. Activities executed by students or teachers through any of these tools or application are logged and stored in the Learning Record Store (LRS). The data objects collected in the LRS therefore capture all the learning activities in the Up2U ecosystem. These objects are defined within the project as *Category-1* data. The purpose of the collection is to provide a large, comprehensive and integrated set of data featuring activities in formal and informal learning spaces. This data set is of high interest to learning analytics researchers. The types and formats of the respective data objects are not yet defined as the ecosystem is still under development. Protocol-wise, the eXperience API (<http://tincanapi.com/overview/>) will be used to collect the data with the respective data types still to be defined. The expected size of the data set is also unknown as of today, but it is expected that the overall size of the raw activity data will be in the range of Terabytes. Any further details and changes related to *Category-1* data will be reported in future version of this document.

Up2U conducts surveys to gather information about the ICT situation of schools as well as their needs, and also to get opinions of responsible people at schools regarding open educational resources, security and privacy, IPR, and the upcoming GDPR. This data is defined as *Category-2*. The purpose of collecting this data is to design, implement, and deploy an Up2U infrastructure that fulfils the needs of the schools. Furthermore, this data offers a unique opportunity to decision makers, governments, and the schools itself to learn about and understand the situation at European schools from the perspective of Up2U. The data is collected through Google Forms (<https://www.google.com/forms/about/>) and stored as Microsoft Excel files. The overall size of the survey data is in the range of Megabytes. Any further details and changes related to *Category-2* data will be reported in future version of this document. Furthermore, new *Categories* of data, which are collected or generated within the project, will be added to the data management plan.

As of today, Up2U has not re-used any particular data objects from other third parties (but software, tools, applications, and infrastructures), but it will in future if appropriate. Any re-use will be reported in future versions of this document.

What is the purpose of the data collection/generation and its relation to the objectives of the project?

What types and formats of data will the project generate/collect?

Will you re-use any existing data and how?

What is the origin of the data?

What is the expected size of the data?

To whom might it be useful ('data utility')?

2. FAIR data

Regarding means to find, access, make interoperable, and re-use the data collected by Up2U, the two categories are handled differently. *Category-1* data objects are of interest for researchers and therefore have to be made findable, accessible and interoperable, and their re-use will be fostered. *Category-2* data objects are used internally in the Up2U project and mainly the knowledge derived from them will be published. Any sharing of the basic *Category-2* data objects is not yet planned.

Any future updates of the FAIR handling of *Category-1* and *Category-2* data objects will be reported in this document.

2.1 Making data findable, including provisions for metadata

2.1.1 Category-1 data objects

As the Category-1 data objects are not fully specified regarding their type and format, it is too early to provide information regarding "making data findable". Up2U will, however, make sure that suitable identifier and metadata standards or best practices will be applied.

2.1.2 Category-2 data objects

The data objects of this category are not shared.

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

What naming conventions do you follow?

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

2.2 Making data openly accessible

2.1.1 Category-1 data objects

Category-1 data will be made accessible. As the Category-1 data objects are not fully specified regarding their type and format, it is too early to provide information regarding the accessibility. Up2U will take care that a suitable repository is chosen depending on the requirements of the research community. It is envisaged that no particular software will be necessary to access the data. Up2U will make sure that the published data will not contain any personal data. It is therefore currently not planned to implement a data access committee or restrict the data access as open access is preferred.

2.1.2 Category-2 data objects

The data objects of this category are not shared. They contain personal data about the people answering the surveys as well as their opinions. Furthermore, these data objects are generated to shape the Up2U ecosystem and are not research data to be shared per se. As stated above, anonymous statistics derived from these objects might be of value to certain stakeholder, but they are results of an analysis and not the data objects themselves.

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

What methods or software tools are needed to access the data?

Is documentation about the software needed to access the data included?

Is it possible to include the relevant software (e.g. in open source code)?

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

Have you explored appropriate arrangements with the identified repository?

If there are restrictions on use, how will access be provided?

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine readable license)?

How will the identity of the person accessing the data be ascertained?

2.3 Making data interoperable

2.1.1 Category-1 data objects

As the Category-1 data objects are not fully specified regarding their type and format, it is too early to provide information regarding their interoperability. Up2U will, however, make sure that suitable standards are chosen wherever possible to ease interoperability.

2.1.2 Category-2 data objects

The data objects of this category are not shared.

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

2.4 Increase data re-use (through clarifying licences)

2.1.1 Category-1 data objects

As the Category-1 data objects are not fully specified regarding their type and format, it is too early to provide information regarding their re-use. Up2U will most likely license the data under a Creative Commons license and will not make any restrictions regarding the duration of their re-use. Further details have to be specified during the coming project months.

2.1.2 Category-2 data objects

The data objects of this category are not shared.

How will the data be licensed to permit the widest re-use possible?

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

How long is it intended that the data remains re-usable?

Are data quality assurance processes described?

3. Allocation of resources

The costs for making data FAIR in Up2U depend mainly on the yet to be specified details of the data objects. This cost will be, for the lifetime of the project, covered by the Up2U consortium. The data management will be governed by WP 6. Regarding long-term preservation, the resources will be determined during the project life-time based on the cost.

What are the costs for making data FAIR in your project?

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Who will be responsible for data management in your project?

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

4. Data security

Partner GWDG will make sure that data security and data protection is taking care of with respect to the data objects collected in the project. GWDG is a data centre and has the respective expertise handling inter alia data from medical and sociological research. With respect to the decision of which repository to choose, certification and data security will be taken into consideration.

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

5. Ethical aspects

The ethical issues identified with respect to Category-1 and Category-2 data have been set out in the confidential Ethic Deliverables: D9.1, D9.2 and D9.3.

Up2U will inform its users regarding the data collection details through the Learning Management System.

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

6. Other issues

No further issues have been identified so far.

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

SUMMARY TABLE 1

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed
1. Data summary	<ul style="list-style-type: none"> • State the purpose of the data collection/generation • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data • State the expected size of the data (if known) • Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	<ul style="list-style-type: none"> • Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
2.2 Making data openly accessible	<ul style="list-style-type: none"> • Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	<ul style="list-style-type: none"> • Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow interdisciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"> • Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
3. Allocation of resources	<ul style="list-style-type: none"> • Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation
4. Data security	<ul style="list-style-type: none"> • Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	<ul style="list-style-type: none"> • To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	<ul style="list-style-type: none"> • Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

