EVO NANO

Project:

EVO-NANO

Grant Agreement (GA) No. 800983

"EVOLVABLE PLATFORM FOR PROGRAMMABLE NANOPARTICLE-BASED CANCER THERAPIES"

Call: H2020-FETOPEN-2016-2017
Topic: FETOPEN-01-2016-2017
Type of action: Research and Innovation action (RIA)
Start date of project: 01/10/2018 Duration: 36 months

D5.2: DATA MANAGEMENT PLAN

DELIVERABLE FACTSHEET

Project title | Acronym | Number		Evolvable platform for programmable nanoparticle-based cancer therapies | EVO-NANO | 800983	
Due Date:	31/03/2019	**Date of submission:**	14/01/2019
Month of Project	M06	**Month of submission:**	M04
Title of deliverable:	D5.2 – Data management plan	**Work Package:**	WP5 – Project management
Dissemination level:	PU	**Version/Status**	1.0 / final
Deliverable leader (Name|Organisation)	Andrew Adamatzky |UWE BRISTOL	**Editor(s)**	Andrew Adamatzky, Igor Balaz
Contribution of partners	Andrew Adamatzky and Igor Balaz compiled report, all other members gave their suggestions		
Final review and approval	All partners		
Keywords	Data, plan, zenodo, management		
Abstract	First version of the Data management plan for the project EVO-NANO. It is based on FAIR (Findable, Accessible, Interoperable, Reusable) data management and is compiled following structure of the Horizon 2020 FAIR Data Management Plan (DMP) Template.		
Document change history			
Date	**Authors**	**Description**	

08/10/2018	Andrew Adamatzky	Wrote first version of the document
10/10/2018	Igor Balaz	Added description for parts of Sections 1 and 2
29/11/2018	Andrew Adamatzky and Igor Balaz	Compiled final version of the document following suggestion of all other Consortium members

CONSORTIUM

	Name	Short Name	Country
1.	Univerzitet u Novom Sadu, Poljoprivredni fakultet Novi Sad	UNSPF	Serbia
2.	University of Bristol	UNIVBRIS	United Kingdom
3.	University of the West of England, Bristol	UWE BRISTOL	United Kingdom
4.	Abo Akademi	AAU	Finland
5.	Fundacion IMDEA Nanociencia	IMDEA NANO	Spain
6.	Prochimia Surfaces SP. ZO.O.	PCS	Poland
7.	Fundacio Hospital Universitari Vall D'Hebron – Institut de Recerca	VHIR	Spain

EXECUTIVE SUMMARY

The Data Management Plan of the EVO-NANO project adheres to the following principles: (i) The publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner. (ii) Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. (iii) Data with acknowledged long-term value should be preserved and remain accessible and usable for future research. (iv) To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. (v) Published results should always include information on how to access the supporting data. (vi) To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process. (vii) To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils. (viii) In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed. (ix) It is appropriate to use public funds to support the management and sharing of publicly-funded research data. (x) To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

TABLE OF CONTENTS

LIST OF FIGURES

none

LIST OF TABLES

none

LIST OF ABBREVIATIONS

Abbreviation	Description
DDS	Drug Delivery System
DMP	Data Management Plan
DOI	Digital Object Identifier
EU	European Commission
FAIR	Findable, Accessible, Interoperable, Reusable
NP	Nanoparticle

1. DATA SUMMARY

What is the purpose of the data collection/generation and its relation to the objectives of the EVO-NANO?

Long term vision of EVO-NANO project is to create an integrated platform for the artificial evolution and validation of novel DDS for cancer treatment using NP. Within EVO-NANO we defined four specific objectives:

Objective 1: To develop a new class of open-ended evolutionary algorithms to creatively assess different cancer scenarios and autonomously engineer effective NP-based solutions to them in a novel way.

Objective 2: To implement a computational platform for the autonomous generation of new strategies for targeting CSC surface receptors using functionalized NPs. In its final form our model will simulate all the main aspects of NP dynamics: their travel via blood streams, extravasation, tumour penetration and endocytosis.

Objective 3: To streamline synthesis of functionalized NPs suggested by the computational platform.

Objective 4: To develop an integrated platform for validation of efficacy of the artificially evolved nanoparticle designs. It will be composed of (i) tumour microenvironments on microfluidic chips that will mimic major physiological barriers for NP tumour delivery and (ii) in vivo pre-clinical tests.

Reaching any of these objectives will require generation and/or collection of specific data:

- Collection of available data on physico-chemical NP properties, nonspecific chemical interactions of functionalized NPs in the bloodstream, extravasation of NPs, interaction of NPs with tumour cells, behaviour of NPs within tumour cells (Objective 1);
- Source code data (Objective 1 & 2) ;
- Simulation output files (Objective 2) ;
- Analysis of simulation output files (Objective 2) ;
- Characterization of synthesized NPs (Objective 3) ;

- Results of *in vitro* and *in vivo* tests (Objective 4).

In summary, data collection/generation follows different procedures for each of the proposed objectives and within EVO-NANO we will create at least 7 separate datasets. Since reusability of datasets between different research groups is essential for the success of the project, development of DMP is equally important.

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

- Data from computer models, represented as text, binary or graphics files and videos.
- Data from clinical assessment of the treatment.
- Measurements in biological matrices/tissues.
- Molecular and chemical data on nanoparticles developed, including core and coating .
- Exposure data: internal biomarkers of exposure.

The Open Research Data Pilot applies to two types of data:

1) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible.

2) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan–that is, according to the individual judgement by each project group.

According to the "Guidelines on Data Management in Horizon 2020" (2015) the DMP describes the handling of numerical datasets processed or collected during EVO-NANO lifetime. The DMP include clear descriptions and rationale for the access regimes that are foreseen for collected data sets. Thus the DMP leaves explicitly open the handling, use and curation of products like tools, software and written documents, which could also be subsumed under the generic term "data"; we restrict the focus of our DMP to numerical data products like produced model data or observation data.

Formats of the data:

- Data and metadata will be requested, stored and transferred (across partners and in EVO-NANO) in a comma separated values (CSV) format.
- To facilitate the data exchange, MS Excel compatible files including comma separated and .xls(x) format will be also accepted.
- For statistical purposes, other formats include .sas7bdat (SAS), .RData (R), .SAV (SPSS), .mat (matlab).
- Where applicable data formats may be migrated when new technologies become available and are proved robust enough to ensure digital continuity and continued availability of data

We will follow the following guidelines:

- Guidelines on Data Management in Horizon 2020, Version 2.0, 30 October 2015: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 2.0, 30 October 2015: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- Webpage of European Commission regarding Open Access: http://ec.europa.eu/research/science-society/open_access

Will you re-use any existing data and how?

To develop NP simulations we will re-use available published data on:

- physico-chemical NP properties,
- nonspecific chemical interactions of functionalized NPs in the bloodstream,
- extravasation of NPs,
- interaction of NPs with tumour cells,
- behaviour of NPs within tumour cells,

and use them to define model boundary conditions.

What is the origin of the data?

- Written source code;
- Published scientific articles;
- Outputs of *in silico* experiments;
- Outputs of tools for mathematical analysis;
- Experimental *in vivo* and *in vitro* test;
- Characterization results of synthesized NPs.

What is the expected size of the data?

- Numerical data related to optimisation c. 250 Gb per project's lifetime;
- Full history of the 3D models of nano-particle swarm: c. 5Tb;
- Videos of the computer models and animation of results: c. 10Tb;

The above are estimates. To be evaluated during the course of the project. The expected size depends on the extend and the nature of the data that are made available.

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

- EVONANO consortium;
- European Commission services and European Agencies;
- EU National Bodies;
- The general public including the broader scientific community
- Manufacture of nanoparticle based treatment
- Clinicians using the nanoparticles

2. FAIR DATA

2. 1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

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)?

Yes. All EVO-NANO generated data are stored at Zenodo repository (https://zenodo.org/). Zenodo closely follows FAIR principles:

- Each uploaded dataset gets DOI;
- Metadata for individual records are retrievable by their identifier using a standardized communications protocol;
- Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it;
- Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation

All data will be discoverable via medata provision. All data will be identifiable and referable via standard identification mechanism (DOI). A unique online naming conventions are adopted. The data will be searchable by keywords. Clear versioning will be in place.

What naming conventions do you follow?

To be able to clearly distinguish and identify data sets, each data set is assigned with a unique name. To design the data set names, we use the following procedure:

FieldIdentifier.CountryCode.PartnerName.DatasetName, where

a. *FieldIdentifier* defines subfield within which data are produced
b. The *CountryCode* part represents the country associated with the dataset using ISO Alpha-3 country codes:
   i. FIN for Finland
   ii. POL for Poland
   iii. SRB for Serbia
   iv. ESP for Spain
   v. GBR for United Kingdom
c. The *PartnerName* part represents the name of the organization associated with the dataset:
   i. UNSPF for Univerzitet u Novom Sadu, Poljoprivredni fakultet Novi Sad
   ii. UNIVBRIS for University of Bristol
   iii. UWEBRISTOL for University of the West of England
   iv. AAU for Abo Akademi
   v. IMDEANANO for Fundacion IMDEA Nanociencia
   vi. PCS for Prochimia Surfaces SP. ZO.O.
   vii. VHIR for Fundacio Hospital Universitari Vall D'Hebron – Institut de Recerca
d. The *DatasetName* represents the full name of the dataset.

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

Yes.The dataset information reported into the metadata fiche will be published in EVO-NANO, where specific filters, based on the metadata elements, will allow to refine the search across datasets (e.g. search dataset by chemical or chemical group, by temporal or spatial coverage of the data, by key words, etc.

Do you provide clear version numbers?

Yes. The versioning management of the data, metadata template and in general the files stored into the Repository will be applied at two levels:

1. Via the naming convention and the use of the date as suffix, indicating the last version of the file uploaded into the Repository;
2. As capabilities of the Zenodo Repository sets up for the project, since the solution supports the simple version control system for the uploaded files

What metadata will be created?

We will adopt a metadata scheme used to describe chemical monitoring data collections, e.g.:

chemical occurrence data collected as result of legal obligations on adhoc or regular basis for reporting /monitoring at European or national levels;

data generated as result of targeted research on the presence of known or unknown chemical substances in specific media in a European country/region.

Metadata is compliant with two European metadata standards, namely:
1. INSPIRE metadata elements for spatial data sets and services (see these elements in the INSPIRE Metadata Regulation: http://data.europa.eu/eli/reg/2008/1205/oj#d1e600-14-1)
2. The "DCAT application profile for European data portals" (DCAT-AP), developed in the framework of the EU ISA Programme.

The European Data Portal is implementing the DCAT-AP as the common vocabulary for harmonising descriptions of datasets harvested from several data portals of 34 countries. The DCAT-AP specification is available at: https://joinup.ec.europa.eu/asset/dcat_application_profile/

2.2. MAKING DATA OPENLY ACCESSIBLE

Which data produced and/or used in the project will be made openly available as the default?

When no embargo period applies and a data package related to a case study has been marked as public, it will be made openly available. Only data gathered by partners outside of the project work plan and protected by IPR, or inside the work plan but containing confidential information (e.g. patent application) will be kept closed until those results are necessary to protect, in accordance to Articles 27, 29, 36, 37 and 39 of the EVO-NANO Grant Agreement number 800983. These principles

will apply to the following inclusive but not an exhaustive list of sets of data produced by the EVO-NANO:

- Material science data on nanoparticles
- Results of computer modelling
- Optimisation analysis and results
- Results of experimental laboratory trials
- Results of clinical studies

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

All generated datasets within EVO-NANO will be uploaded to Zenodo repository (https://zenodo.org/).The data sharing should occur in a timely fashion.This means that the data resulted from the research conducted in the project should become available close to the project results themselves. Furthermore, it is reasonable to expect that the data will be released in waves as they become available or as main findings from waves of the data are published.

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

Since Zenodo stores data as publicly accessible, the only requirement is internet access. With regards to open software, all the data needed to create and maintain the marketplace is being made openly accessible through the GitHub repository, along with the corresponding technical documentation.

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

No documentation is required. Online help provided with existing browser is sufficient.

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

Software will be shared via GitHub, which is directly linked with Zenodo.

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.*

The consortium agreed to deposit the data generated by the project in Zenodo, publications in arXiv, and software in GitHub unless for a specific project there is a subject specific repository that is considered more relevant.

Have you explored appropriate arrangements with the identified repository?

Yes, the arrangement are tested by the Partners in their other projects.

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

There are not restrictions to us. To access data no registration at arXiv, Zenodo or GitHub is required.

Is there a need for a data access committee?

There is no need for a data access committee because sharing of data is agreed straightforwardly.

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

Potential users will find out about the data through publications and the website. Data will be made available on publication of the associated paper and will be made accessible on request, under conditions agreed on a case-by-case basis, and after agreement of the project consortium.

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

The identity of the person accessing the date will not be ascentain because their access is anonymous.

2.3. MAKING DATA INTEROPERABLE

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)?

All data produced will have transparent formats: publications in PDF, compute code in Phython/C++/Processing/Java, results of numerical simulation in CSV, animations and videos in AVI/MP4.

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

Other types of data have been registered following internal codifications, clearly specified within the file.

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?

Not applicable.

2.4. INCREASE DATA RE-USE (THROUGH CLARIFYING LICENCES)

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

The deliverables associated to the dataset are licensed through an All rights reserved license as they are working papers not intended to be re-used. Nevertheless the database should be shared as a possible reusable dataset. For this reason, when deposited to the repository, an Attribution-NonCommercial license (by-nc) will be requested. The data is currently available for re-use from the project website and will also be findable and reusable through the final depositing repository (the institutional one or Zenodo) and from OpenAire, the latest by the end of the project.

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.*

The data will remain re-usable after the end of the project by anyone interested in it, with no access or time restrictions.

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.*

Each archived data set will have its own permanent repository ID and will be easily accessible. We expect most of the data generated to be made available without restrictions and only data sets subject to IPR and confidentiality issues will be restricted. Where this is going to be the case, agreements will be made based on the individual data sets. Requests for the use of the data by externals will be approved by the project consortium.

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

Data and metadata will be retained for the lifetime of the Zenodo repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.

Are data quality assurance processes described?

The data quality is ensured by different measures. These include validation of the sample, replication and comparison with results of similar studies and control of systematic distortion.

3. ALLOCATION OF RESOURCES

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

Exact costs estimated will be known and adjusted dynamically during the project's lifetime.

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).*

The costs for depositing the dataset with the project, and subsequent resources required to make the dataset publicly available have been included within specific WPs within the project.

Who will be responsible for data management in your project?

The project coordinator has the ultimate responsibility for the data management in the project and so, for the Marketplace platform management.

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)?

Due to the data being shared via public repositories, the preservation beyond lifetime of the project does not involve any costs.

4. DATA SECURITY

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

Due to the data volume, zenodo also hold a copy of their own processed data, effectively acting as a second distributed database and additional backup. Locally, within each partner, all data will be stored at backup external hard-disks.

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

The digital signature of the whole dataset, or the storage of the dataset in a git repository could provide support for the correct duplication and preservation. In addition zenodo operates with 12-hourly backup cycle with one backup sent to tape storage once a week.

5. ETHICAL ASPECTS

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).*

The ethical aspects related to the personal data collected in this dataset are addressed in the Ethics Requirements document of the original proposal. Regarding the protection of personal data of the research participants, the Consortium will meet the following conditions:

- To submit to the REA the copies of ethical approvals for the collection of personal data by each of the competent University Data Protection Officers or National Data Protection authorities.
- To justify (if necessary) the collection and/or processing of personal sensitive data.
- To follow and accomplish the national and EU legislation on the procedures that will be implemented for data collection, storage, protection, retention and destruction.

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

Not applicable in this project.

6. OTHER ISSUES

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

Through the use of the institutional repository, we are also following these procedures for data management Partner countries, including the Partner countries Research Council's common principles on data policy provide an overarching framework for individual Research Council policies on data policy.