WIN for publication of digital objects in Knowledge Junction

What is Knowledge Junction?

The Knowledge Junction is a curated community in the Zenodo repository for research sharing. Zenodo is an open source product, built on the foundation of the CERN Invenio digital library. It is used by the OpenAIRE project, and was commissioned by the EC to support their nascent Open Data Policy by providing a catch-all repository for EC funded research. This community enables the exchange of evidence and supporting materials used in food and feed safety risk assessments, with the goal of improving transparency, reproducibility and evidence reuse. Submissions of evidence (reports, datasets, images, videos, laboratory outputs, etc.) and supporting materials generated during risk assessment (software, tools, models, code, protocols, appraisal schemes, FAQs etc.) in all file formats are welcome from any registered user of Zenodo. Member State organisations, EEA/EFTA countries and Pre-Accession countries are encouraged to upload, through national representatives, all information relevant to risk assessment activities including risk assessment mandates, outputs (opinions, reports, statements, guidance documents), national workplans related to risk assessment, relevant technical reports, any other relevant document, data and software.

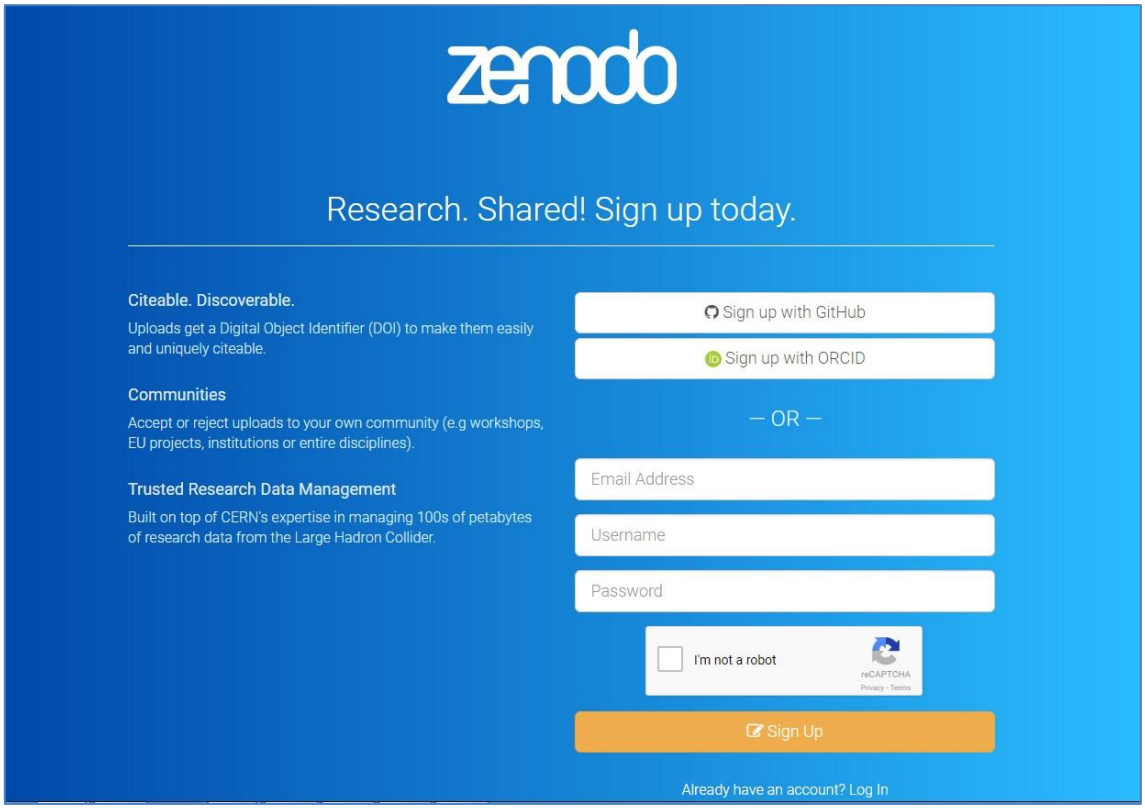
A DOI is issued for every publication

Each submission is assigned a unique and persistent digital object identifier (DOI) that allows easy retrieval and persistence of published items. Each publication requires metadata acknowledging authors and contributors and has a license specifying the conditions of use. In this way the reuse of any object deposited can be tracked by the provider. The community is intended for uploading digital files that have not been published elsewhere with a Digital Object Identifier (DOI). However, if the material is not subject to copyright restrictions, the file can be uploaded into the community and the existing DOI should be used. All content submitted is viewed by the community curator for acceptance and publication and high quality metadata is essential. The content of this repository can be used by EFSA's Panels and Working Groups and any other interested parties when preparing new risk assessments. The Knowledge Junction enables broad dissemination of evidence, collaboration between countries and national organisations, avoids duplication of activities and increases visibility of countries' activities in the area of food and feed safety risk assessment. The platform is extremely user-friendly and intuitive with an advanced end-user interface and is visually appealing. The following screenshots will guide you through the process of registration in the Platform and publication of digital objects in the community.

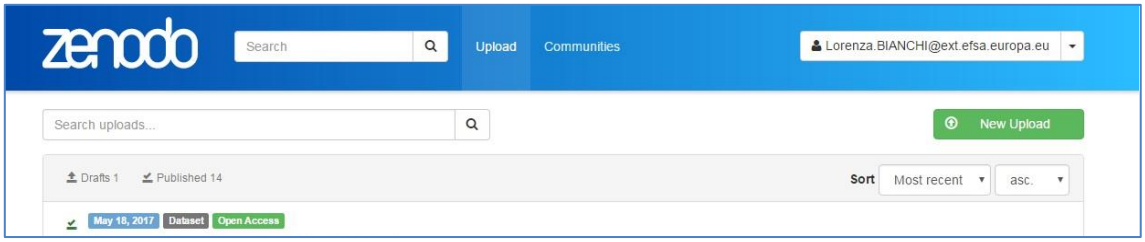
Step-by-step guide to publish a digital object in the Knowledge Junction

Start the publication

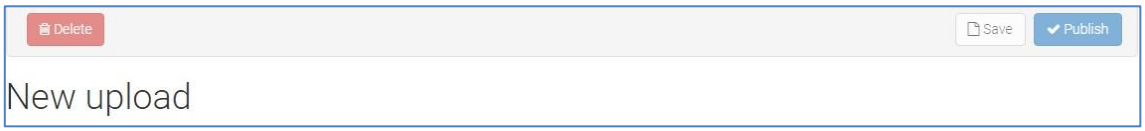
To publish a digital object in the Knowledge Junction, you need to sign-up first on the Zenodo Platform. If you already have an ORCID you can use this to login.



Once you have created your **own account**, you can proceed with by clicking on the **New Upload** button.

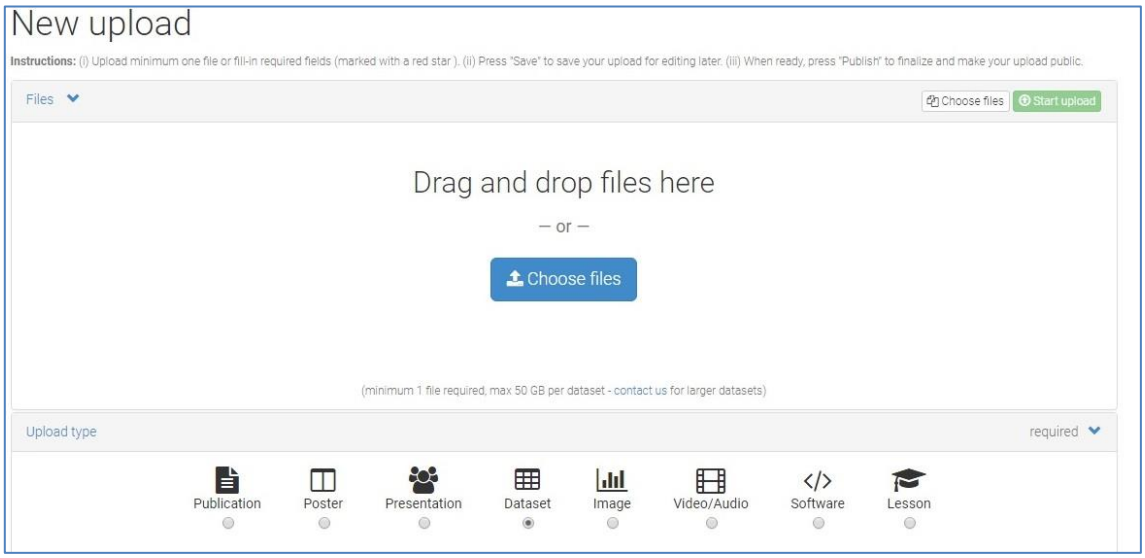


From the upload page, you can start the publication of your file/s by entering the required metadata. We recommend that you save your information after each modification by clicking on the save button so not to lose it in case of connection problems. Please bear in mind that you can only start saving after you have filled in the mandatory fields.



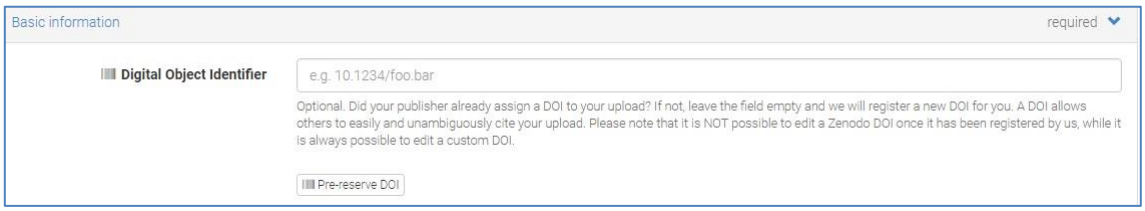
Upload a digital object

First, you must select the electronic file/s that you want to publish and click on the **Start upload** button. The second step is to select the **upload type** (in the example we have selected the dataset type). All file formats are allowed.



Zenodo automatically generates a DOI (digital object identifier) so this can be ignored unless your file/s already have a DOI, in which case then you should enter the existing DOI (please do not upload objects which have been previously published and are subject to copyright restrictions).

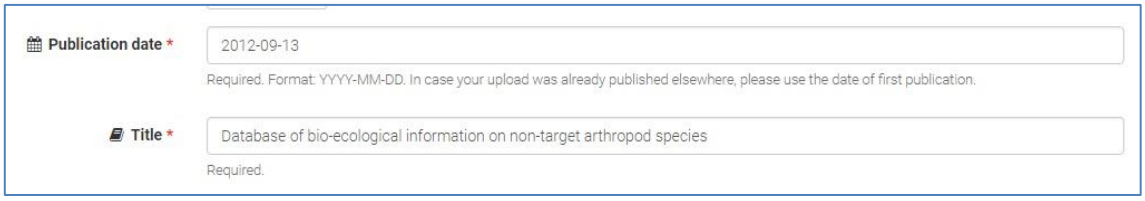
It is also possible to pre-reserve a DOI prior to publication by clicking on the Pre-reserve DOI button.



Add metadata

You can now proceed with the provision of **mandatory metadata**:

- **Publication date**: the day of submission is the default date of publication. If the item was already published elsewhere, please use the original date of publication.
- **Title:** you must provide a short name for the file/s to be submitted.

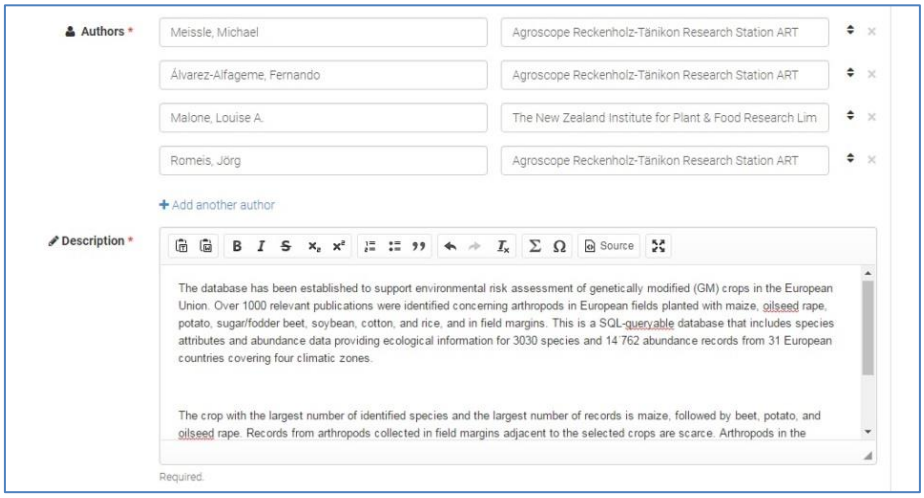


- **Authors:** The author is the entity primarily responsible for making the resource. As a minimum, you must provide the name of the organisation/affiliation that created the dataset. However, since the repository is designed to foster collaboration, the names of the authors are preferred as this allows users to contact them.

  Names should be reported as Surname, Forename. For the affiliation, the EU Food Safety Almanac can be used as a source for the confirmation of the organisation names.

  EFSA authorship principles are recommended.

- **Description:** you must provide a description of the file/s which can be used by anyone viewing or downloading your files to understand the content. Useful information can include the objective or purpose for the creation of the file/s, the scope (geographical, temporal, domain, etc.), the methodology or legislative requirements applied, assumptions, uncertainties, key conclusions or any other information that may influence the re-use of the file/s you are sharing.



- **Keywords:** they will help users of the repository to retrieve data that are relevant for their area of research. Use always lower case except for scientific terminology.

- **Additional notes**: This field is used to stimulate collaboration between community members working on similar topics. As a minimum, please indicate:

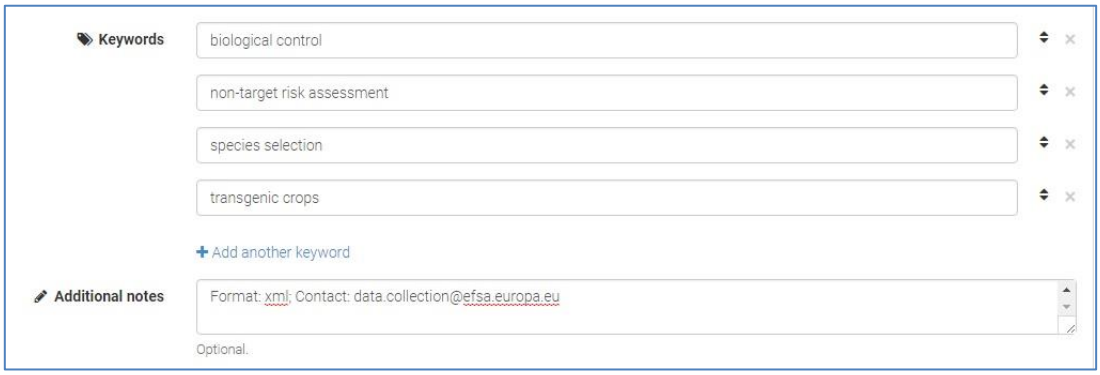
  Submitting country: for Country Code use ISO country standard.

  Language: (original language of the document): for Language Code use ISO country standard.

  Contact: please insert the email of the person submitting the digital object.

  Other information relevant for collaboration: for example institution, format or software required for executing the file in the case of models.

  All fields should be separated by semi-colon, and format and order of variables must be respected. Example: "*IT; en; XLSX; abc@mail.it*".



Define permissions

You can continue the publication process by selecting the access rights and license.

- **Access Right**: the default is Open Access.
- **License**: recommended options for licensing are:

  Creative Commons Attribution 4.0

  Share: copy and redistribute the material in any medium or format;

  Adapt: remix, transform, and build upon the material for any purpose, even commercially;

  Attribution: you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

  Creative Commons Attribution No Derivatives

  Share: copy and redistribute the material in any medium or format for any purpose, even commercially;

  Attribution: you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use;

  NoDerivatives: if you remix, transform, or build upon the material; you may not distribute the modified material.

Creative Commons Attribution – NonCommercial - NoDerivatives

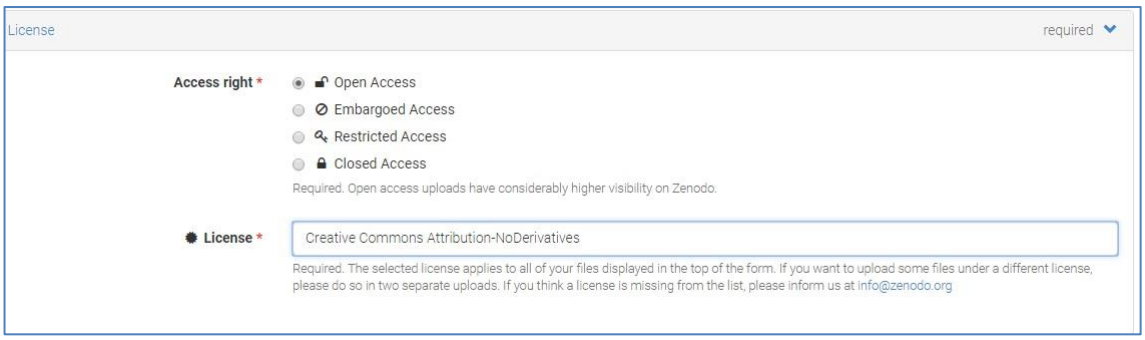
Share: copy and redistribute the material in any medium or format. The licensor cannot revoke these freedoms as long as you follow the license terms;

Attribution: you must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use;

NonCommercial: you may not use the material for commercial purposes;

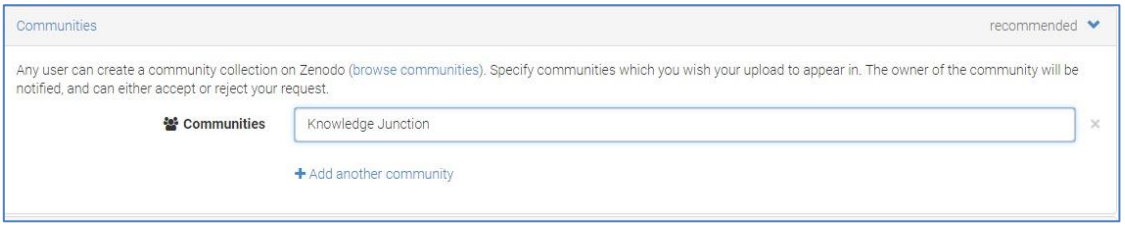
NoDerivatives: if you remix, transform, or build upon the material; you may not distribute the modified material;

No additional restrictions: you may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

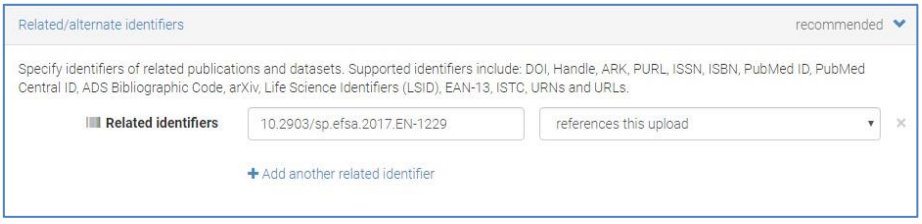


Publish a digital object

- **Communities:** to publish your digital object in the community created by EFSA, you must type and select Knowledge Junction in the specific field (bear in mind that the system is case-sensitive and use initial capital letters).



- **Related links**: this can be used to link to similar or related published items. This can also help other users to learn more about your publication and how it can be re-used appropriately. For example, if you are publishing a dataset which is linked to a report with a DOI or ISSN, then provide the link to the published report.



To continue your publication process, you can provide information on the following optional elements, when appropriate. Please bear in mind the following **definitions**:

- **Contributors:** people listed in the acknowledgements. The process used for authors should also be applied to contributors.
- **Reference**: any references cited under the *Description* should be included here. Additional references may be included at the discretion of the submitter.

  However, EFSA outputs and supporting publications should be linked using the ***related identifiers*** field rather than cited as references, unless no DOI or ISSN exists.

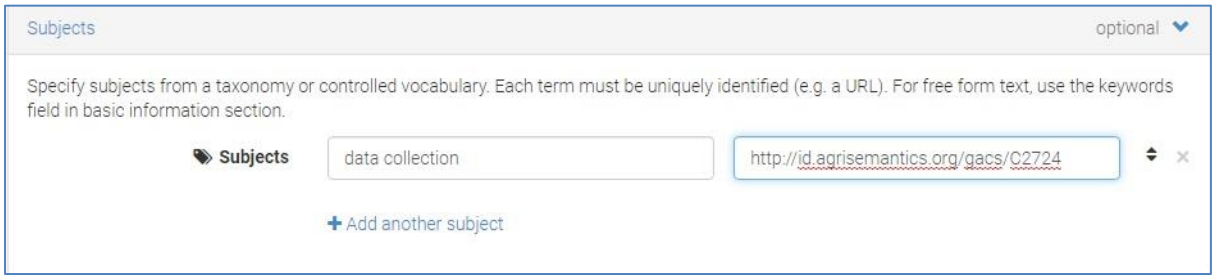
Complete your publication

To complete the publication you are strongly recommended to include subjects:

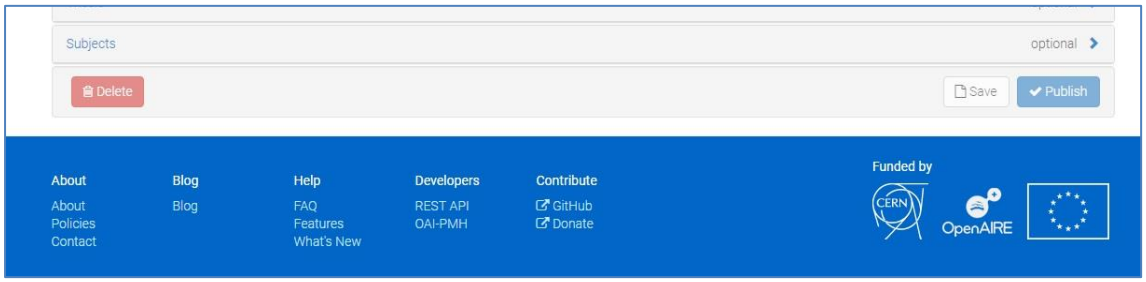
- **Subjects:** each item must be classified using suitable terms selected from the Global Agricultural Concept Scheme.

  The Global Agricultural Concept Scheme (GACS) is a hub for concepts related to agriculture, in multiple languages, for use in Linked Data. It comprises a core set of terms from the *thesauri* of the Food and Agricultural Organization of the United Nations (FAO), CAB International (CABI), and the National Agricultural Library of the USA (NAL).

  Each subject must be uniquely identified (e.g. URI). For free form text, use the **keywords** field in basic information section.



When you have completed the form, you can publish your digital object by clicking on the **publish** button after saving any modification done.



Versioning

Zenodo support the replacement of existing files using versioning, it is possible to cite a specific version of the record (Specific DOI) and to cite all the versions of the record (Concept DOI). In case you wish to modify the metadata you do not need to create a new version, as you just need to click on the Edit button and update the information.

When you publish an item for the first time, Zenodo registers two DOIs, a DOI representing the specific version of your record and a DOI representing all of the versions of your record. Afterwards, Zenodo registers a DOI for every new version of your upload.

General recommendations

Confirm metadata and publication rights

If the submitter is preparing datasets for an organization (e.g. EFSA Focal Point), contact the authors of the dataset to confirm the metadata and inform them of the planned publication in an open repository.

Datasets must not contain personal data. For instance, personal health-related data are sensitive data and therefore are inappropriate for publication. For the definition of personal data refer to Regulation (EC) No 45/2001 (a new version is entering into force during 2018).

Relevant standards, procedures, legislation and documents:

Dublin Core Metadata Element Set

ISO Standard 15836:2009

DataCite Metadata Schema

Global Agricultural Concept Scheme

Creative commons licenses

EFSA authorship principles

The EU Food Safety Almanac

EFSA Strategy 2020

Knowledge Junction

Assure compliance with SOP 014S

When publishing digital objects linked to EFSA outputs follow SOP 014S "Publishing a scientific output in the EFSA Journal".

Annex A: Instructions to upload specific digital items

Specific instructions related to the publication of protocols

Protocols delineate the methodological approach taken when performing an activity.

It is recommended to provide the following metadata information:

Keywords and Subjects

All protocols must report the following information:

Keyword: methodology; subject: methodology; URL: http://id.agrisemantics.org/gacs/C365

In addition, keywords and subjects should cover the specific topic and the type of protocol.

Below you find some examples of keywords and subjects with related URL from GACS:

- **Specific topic, examples:**

  Keyword: plant health; subject: plant condition; URL: http://id.agrisemantics.org/gacs/C21

  Keyword: animal health; subject: animal health URL: http://id.agrisemantics.org/gacs/C15

  Keyword: pesticides; subject: pesticides URL: http://id.agrisemantics.org/gacs/C284

- **Type of protocol, examples:**

  Keyword: systematic review; subject: systematic reviews; URL: http://id.agrisemantics.org/gacs/C10616

  Keyword: extensive literature search; subject: literature review; URL: http://id.agrisemantics.org/gacs/C5999

  Keyword: assessment; subject: assessment; URL: http://id.agrisemantics.org/gacs/C2288

Note: In case the protocol will be published together with the related scientific output the same keywords of this document have to be added.

Specific instructions related to the publication of models

Keywords and Subjects

Recommended terminology for coded objects (terms without a hyperlink can only be reported as keywords):

Multi-compartment model: describes the way materials or energies are transmitted among the compartments of a system. Each compartment is assumed to be a homogeneous entity within which the entities being modelled are equivalent.

Linear models: A model is linear when each term is either a constant or the product of a parameter and a predictor variable. A linear equation is constructed by adding the results for each term.

Nonlinear models: not linear.

Neural networks: model designed to simulate the behaviour of biological neural networks, as in pattern recognition, language processing, and problem solving, with the goal of self-directed information processing.

Network model: a mathematical model where the data is represented as a network of lines and nodes.

Empirical model: model based on investigations, observations, experiments, or experience, as opposed to theoretical knowledge based on logical or mathematical assumptions.

Mechanistic models: a hypothesized relationship between the variables in the data set where the nature of the relationship is specified in terms of the biological processes that are thought to have given rise to the data.

Statistical models: Statistical formulations or analyses which, when applied to data and found to fit the data, are then used to verify the assumptions and parameters used in the analysis.

Agent based models: agent based modelling can be defined as an essentially decentralized, individual-centric (as opposed to system level) approach to model design. Agents can be people, companies, vehicles, cities, animals, ships, products, plants, etc.

Graphical models: define general message-passing algorithms that implement probabilistic inference efficiently.

Simulation models: A mathematical representation of the essential characteristics of a real-world system or situation, which can be used to predict future behaviour under a variety of different conditions.

Hierarchical models: models in which the parameters of prior distributions are estimated from data rather than using subjective information. Hierarchical models are central to modern Bayesian statistics and allow an objective approach to inference.

Calculation: performs basic arithmetic operations or more complex mathematical operations.

Simulation: the act of imitating the behaviour of some situation or some process by means of something suitably analogous (e.g. Monte Carlo method).

Analytical methods: solution directly solved from observed data.

Numerical methods: is a complete and unambiguous set of procedures for the solution of a problem, together with computable error estimates (see error analysis).

Stochastic models: A mathematical model involving random variable(s) in order to estimate probability distributions of potential outcomes.

Deterministic models: a deterministic system is a system in which no randomness is involved in the development of future states of the system. A deterministic model will thus always produce the same output from a given starting condition or initial state.

Time: model includes a temporal component.

Spatial distribution: model includes a spatial component.

Uncertainty analysis: a detailed examination of the systematic and random errors of a measurement or estimate; an analytical process to provide information regarding the uncertainty.

Sensitivity analysis: a technique used to determine how different values of an independent variable impact a particular dependent variable under a given set of assumptions.

Model validation: the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.

Population: a group of individuals defined by a particular set of characteristics, e.g., organisms of a particular species living in a given region or area.

Groups: a number of individual items or people brought together.

Individual: being an individual or existing as an indivisible whole.

Forecasting: the prediction or projection of the nature of future problems or existing conditions based upon the extrapolation or interpretation of existing scientific data or by the application of scientific methodology.

Decision support systems: a computer-based system that enables management to interrogate the computer system on an ad hoc basis for various kinds of information on the organization and to predict the effect of potential decisions beforehand.

Exposure assessment: the determination or estimation (qualitative or quantitative) of the magnitude, frequency, duration and route of exposure of a population.

Sample size determination: it is the act to determine the sample size required to answer a specific research question.

Maximum likelihood method: is a method of estimating the parameters of a statistical model given observations, by finding the parameter values that maximize the likelihood of making the observations given the parameters.

Approximate Bayesian Computation: data generated by simulation are then reduced to summary statistics, and the sampled parameters are 8 accepted or rejected on the basis of the distance between the simulated and the observed summary statistics

Markov Chain Monte Carlo: an iterative Bayesian statistical technique that generates samples from the posterior distribution. Well-designed MCMC algorithms converge to the posterior distribution, which is independent of the starting position, including variants such as particle MCMC.

Document history	
Document reference	*Version 1.1*
Prepared by	*Lorenza Bianchi*
Reviewed by	*Jane Richardson / Sérgio Potier Rodeia / Cristina da Cruz*
Last date modified	*03/08/2017*