NATIONAL OPEN SCIENCE FESTIVAL

MEET / SHARE / INSPIRE / CARE

PROPOSAL FOR WORKSHOP (110 MIN) OR SESSION (45 MIN)

The program committee invites researchers to design and lead an interactive workshop or session for the National Open Science Festival on 1 September 2022. Please submit your proposal by 13 May 2022. Approved sessions will be announced 1 June 2022.

In the morning, there's space for workshops (duration: 110 minutes, 6 slots). In the afternoon, there will be two rounds of sessions (duration: 45 minutes, 12 slots).

We particularly invite proposals for sessions that are interactive and rooted in local initiatives. Therefore, proposals that encourage the participation of attendees will be favoured over proposals with more than 50% of the time reserved for speakers (unless a convincing reason for this format is given). If you want to simply present an initiative we ask you to apply for the call for the Lunch Marketplace.

We are especially interested in contributions that focus on topics such as:

- Practical skill sessions, for instance about preprints, preregistration, use of tools or platforms for practicing open science across the research cycle;
- Open Access in your domain;
- How to involve citizens in your research;
- How to build team science and how to collaborate with not so obvious others;
- Open/FAIR Software;
- Putting FAIR and open data into practice;
- Setting up and maintaining open communities;
- Working-sprints with peers on a specific open topic;
- Network meetings (even if the network has not been formalised yet);
- Proposals that look beyond disciplinary and National boundaries;
- Societal engagement

Please do not feel limited to only these topics.

Proposal length should not exceed 500 words. English is the official language of the Festival.

Names of organiser(s):
Erik-Jan van Kesteren, ODISSEI Social Data Science team.

Names of moderators / speakers if applicable;
Erik-Jan van Kesteren (UU, dept. methods & statistics and ODISSEI SoDa team) Raoul Schram (UU, research IT and ODISSEI SoDa team) Thom Volker (UU PhD student at dept. methods & statistics)

NATIONAL OPEN SCIENCE FESTIVAL

MEET / SHARE / INSPIRE / CARE

please indicate if you want to organize a workshop or session:
• Workshop (110 min)

Short description that summarizes the workshop / session, its rationale and relevance for the participants of the National Open Science Festival;
Open data is one of the pillars of open science. However, there are often barriers in the way of making research data openly available, relating to consent, privacy, or organisational boundaries. In such cases, synthetic data is an excellent solution: the real data is kept secret, but a "fake" version of the data is available. The promise of the synthetic dataset is that others can then investigate the data structure, rerun scripts, use the data in educational materials, or even run a completely different analysis on their own. But how do you generate synthetic data? In this session, we will introduce the field of synthetic data generation and apply several tools to generate synthetic versions of datasets, with various level of utility and privacy. We will be paying extra attention to practical issues such as missing values, data types, and disclosure control. Participants can either use a provided example dataset or they can bring their own data!

Format and structure;
The format will be a workshop-style session: short lectures combined with hands-on sessions - General introduction (15 minutes) - Generating high-privacy synthetic data from metadata (15 minutes) - Hands-on session 1 (30 minutes) - Generating high-utility synthetic data from real data (15 minutes) - Hands-on session 2 (30 minutes) - Closing (5 minutes)

Target audience;
Any researchers interested in sharing their data as part of their open science efforts

Preferred outcomes for participants of the workshop / session;

NATIONAL OPEN SCIENCE FESTIVAL

MEET / SHARE / INSPIRE / CARE

Participants will: - Have an idea how synthetic data may be of use in the framework of open science - Know what the advantages and limitations are of synthetic data generation techniques (the privacy-utility tradeoff) - Be able to create synthetic versions of their data using two different methods

Requirements for the workshop / session (screens, set up of tables, flipcharts, etc.)
Participants should bring their own laptops. We will likely be using both R and python, and we will try to make things run smoothly through e.g. SURF research cloud so that there is very little set-up required Skill requirements for the workshop are that participants have experience with a programming language for scientific use (but participants will be grouped so that even someone with no experience can join & learn) We will need a screen for presentations Tables should be organised In small groups, workshop-style, so that participants can help each other