Работа со складишта за отворени податочни множества

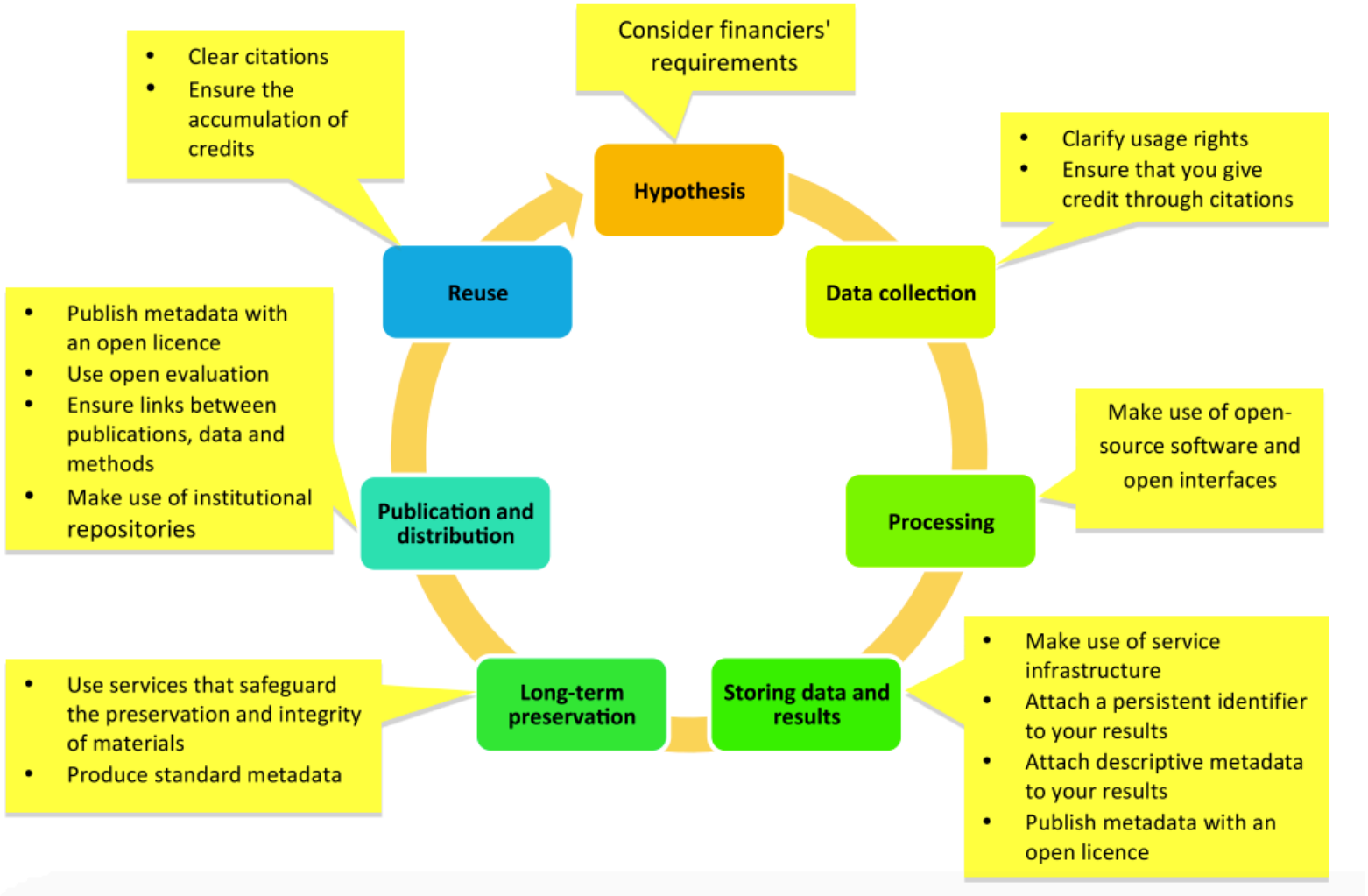
Capacity Building Training
Skopje, 15.10.2021

Anastas Mishev, UKIM

АГЕНДА

- Основи принципи на отворената наука
- Складишта на отворени податочни множества како основа за проверлива и повторлива наука
- Типови на складишта на податоци
- Примери на складишта на отворени податочни множества
- Податочно складиште на репозиториумот на УКИМ
  - Практичен пример за поставување на податочно множество
- Иднината на научните публикации
- Референци

Основни принципи на отворена наука



https://www.fosteropenscience.eu/sites/default/files/pdf/3986.pdf

Складишта на податочни множества

- Складишта на отворени податочни множества се клучни за проверлива и повторлива наука
- Складиште на податоци или податочни множества е генерички термин кој означува архива во која се зачувани податоци за анализа, дистрибуција или известување
- Можат да бидат во строго тесна научна област или мултидисциплинарни
- Квалитет на податоците
  - Начини на проверка на квалитетот
- Опис на податоците
  - клучен за нивно повторно искористување
  - “data curation”
- Неопходен чекор кај одредени издавачи
  - Препорачан или овозможен кај други

За складиштата на податоци и репродуцибилната наука

> - Researchers, on the whole, do not do replication tests as part of their own work, or even as part of the peer review process. In the future, they may be incentives for researchers to do so, and post-publication crowd-sourced peer review in the mold of Wikipedia, as promoted by Edward Curry, may prove to be a successful model.

> - Academic institutions, and their libraries, are increasingly involved in the data management process, but are not involved in replication as a matter of course (note some calls for libraries to take a more active role in this regard).

Типови на складишта на податоци

- Складиштата на податоци, уште наречени и податочни библиотеки или податочни архиви, можат да бидат од различен тип
- Data warehouse – големи складишта кои најчесто складираат податоци од различни извори или сегменти, каде податоците можат да бидат и неповрзани меѓусебно.
- Data lake – големи складишта на неструктурирани податоци опишани со соодветни мета-податоци
- Data marts – подмножества од data warehouse кои содржат податоци релевантни за одредена област
- Складишта на мета податоци - содржат информации за модели на податоци, описи на изворите, значење на податоците итн.
- Data cubes – повеќедимензионални податоци складирани во табели, најчесто содржат временски секвенци од податоци.

European data lakes

Projected figures 2025

530%

increase of global data volume

From 33 zettabytes in 2018 to 175 zettabytes

€829 billion

value of data economy in the EU27

From €301 billion (2.4% of EU GDP) in 2018

10.9 million

data professionals in the EU27

From 5.7 million in 2018

65%

Percentage of EU population with basic digital skills

From 57% in 2018

Примери за складишта на податоци

General repositories

- 4TU.ResearchData
- ANDS contributing repositories
- Dryad Digital Repository
- Figshare
- Harvard Dataverse
- Mendeley Data
- Open Science Framework
- Science Data Bank
- Zenodo
- Code Ocean

Biomedical Sciences

- Eukaryotic Pathogen Database Resources (EuPathDB)
- Influenza Research Database (IRD)
- The Cancer Imaging Archive (TCIA)

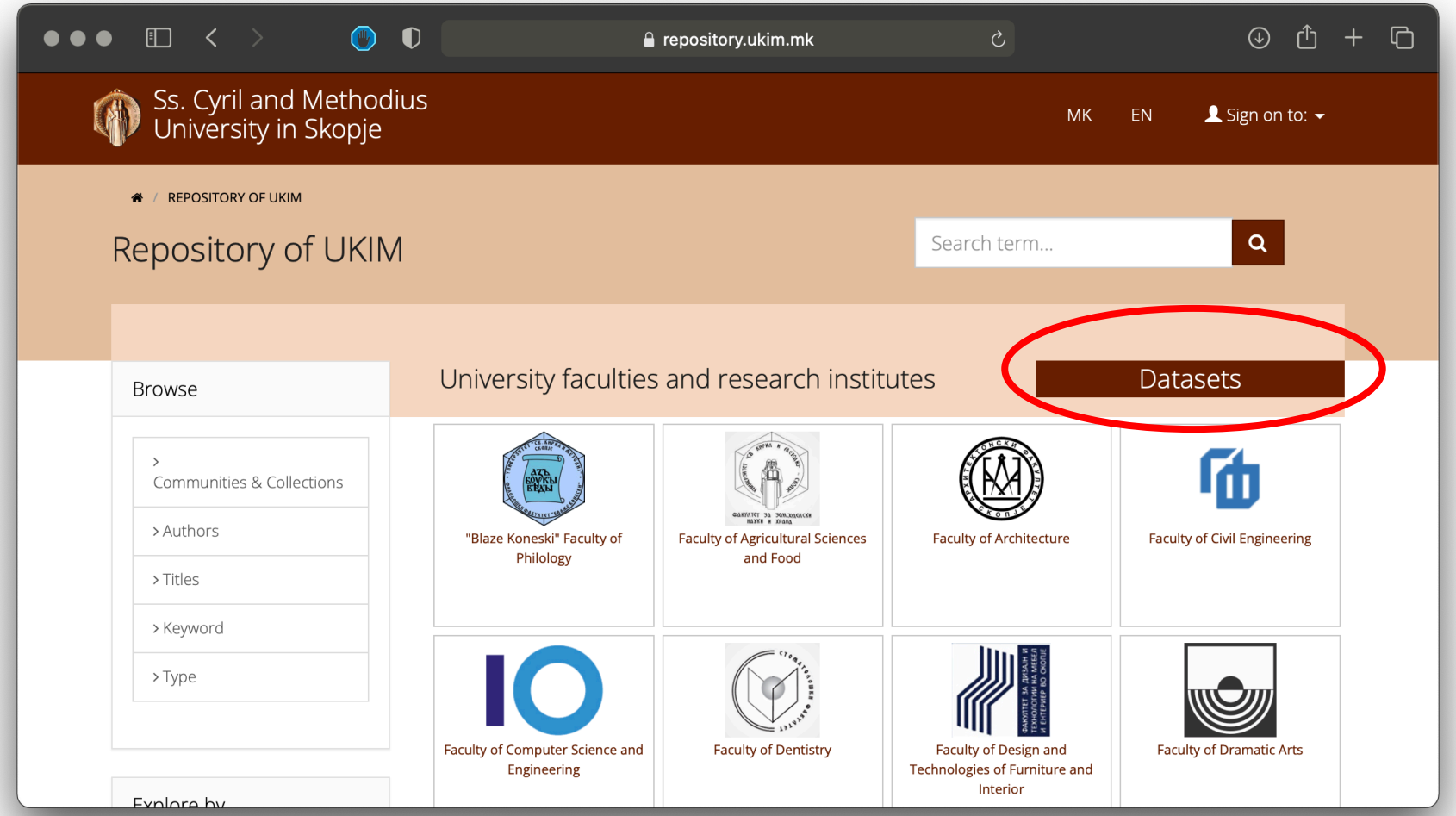
Sequencing

- dbSNP
- dbVar
- Database of Genomic Variants archive (DGVa)
- DNA DataBank of Japan (DDBJ)
- European Nucleotide Archive (ENA)
- GenBank
- MGnify
- miRBase
- NCBI Sequence Read Archive (SRA)
- NCBI Trace Archives
- UniProt

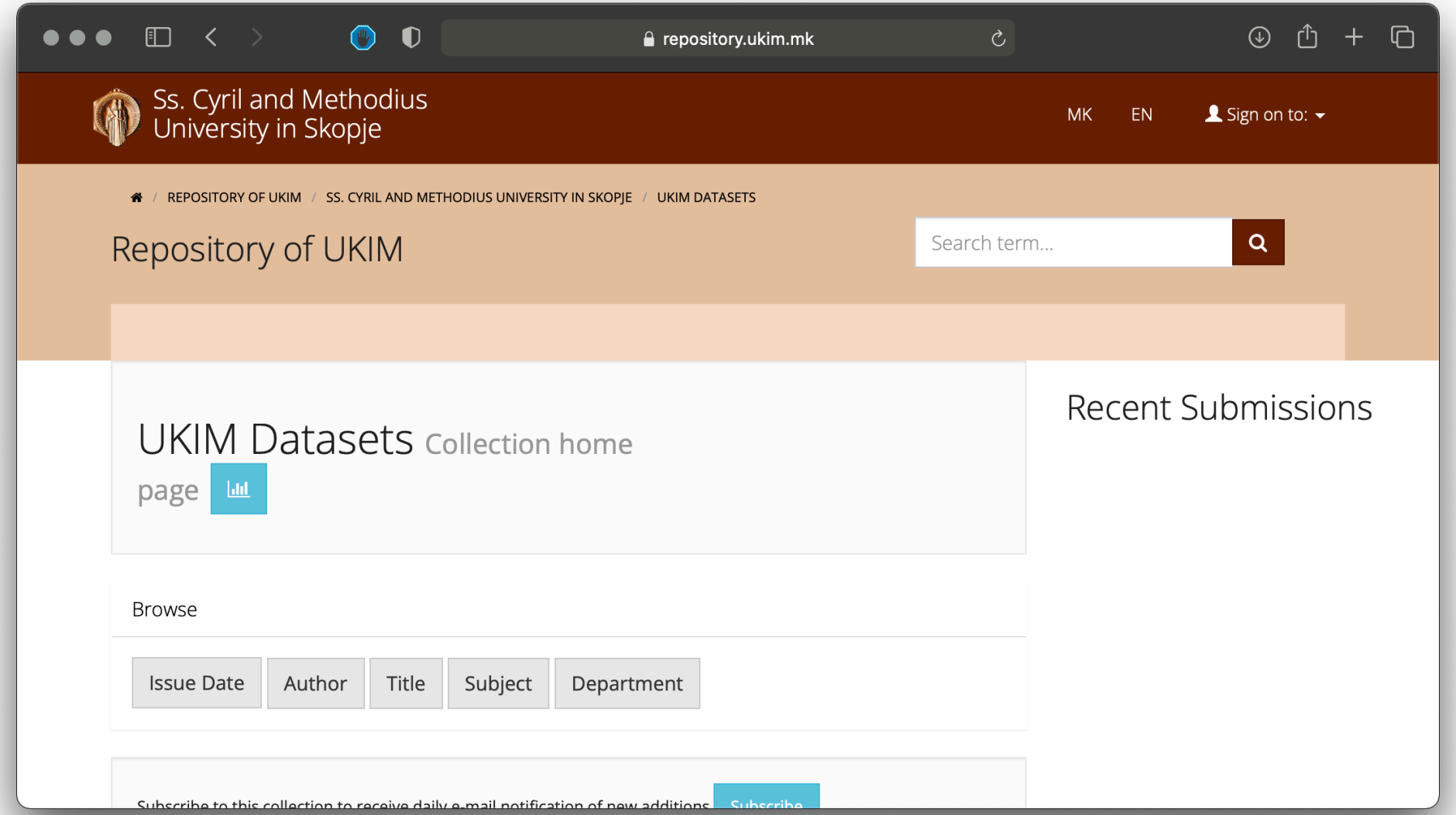
Taxonomic and Species Diversity

- Global Biodiversity Information Facility (GBIF)
- MycoBank
- NCBI Taxonomy
- Integrated Taxonomic Information System (ITIS)

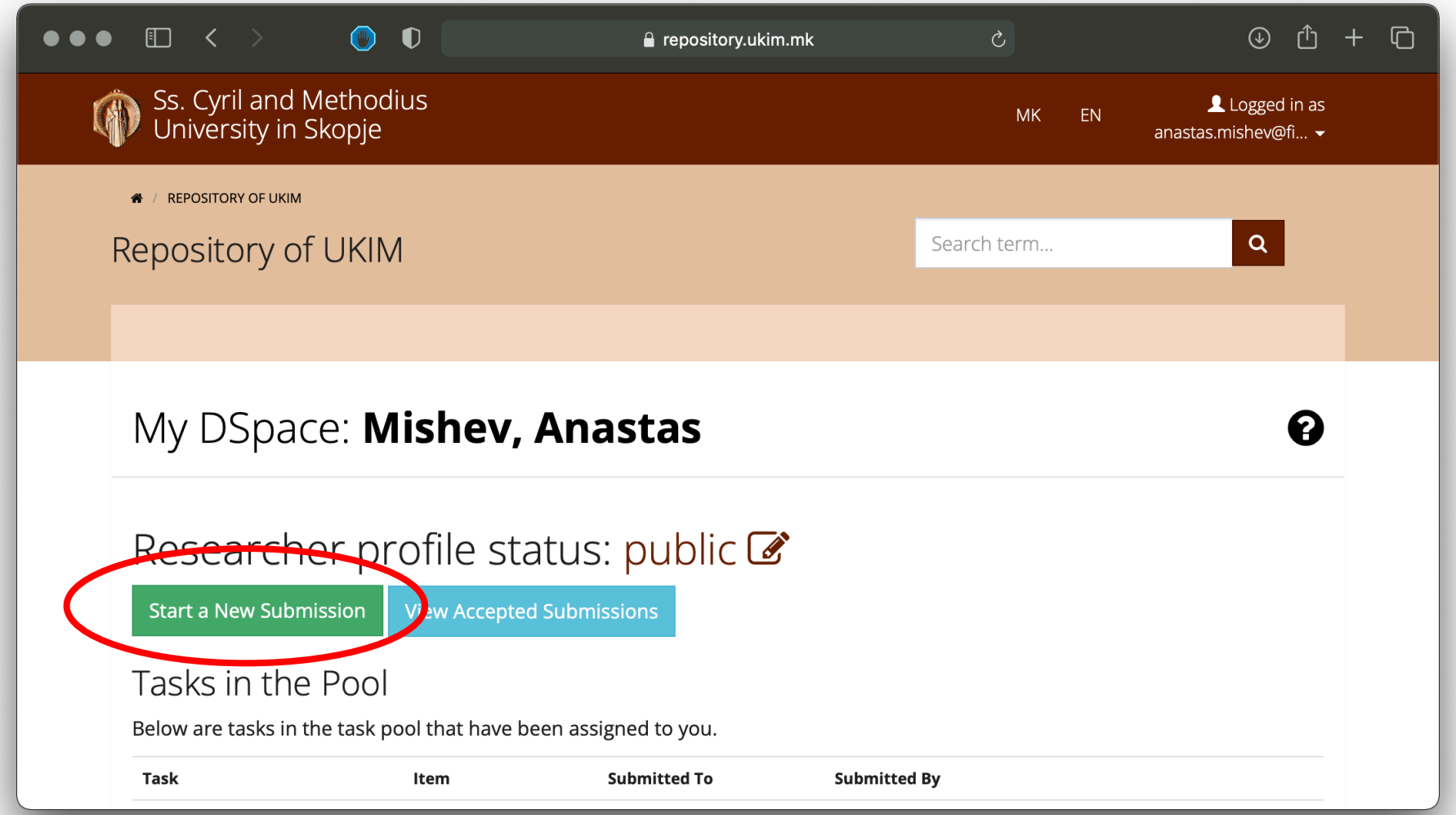
Податочно складиште на репозиториумот на УКИМ



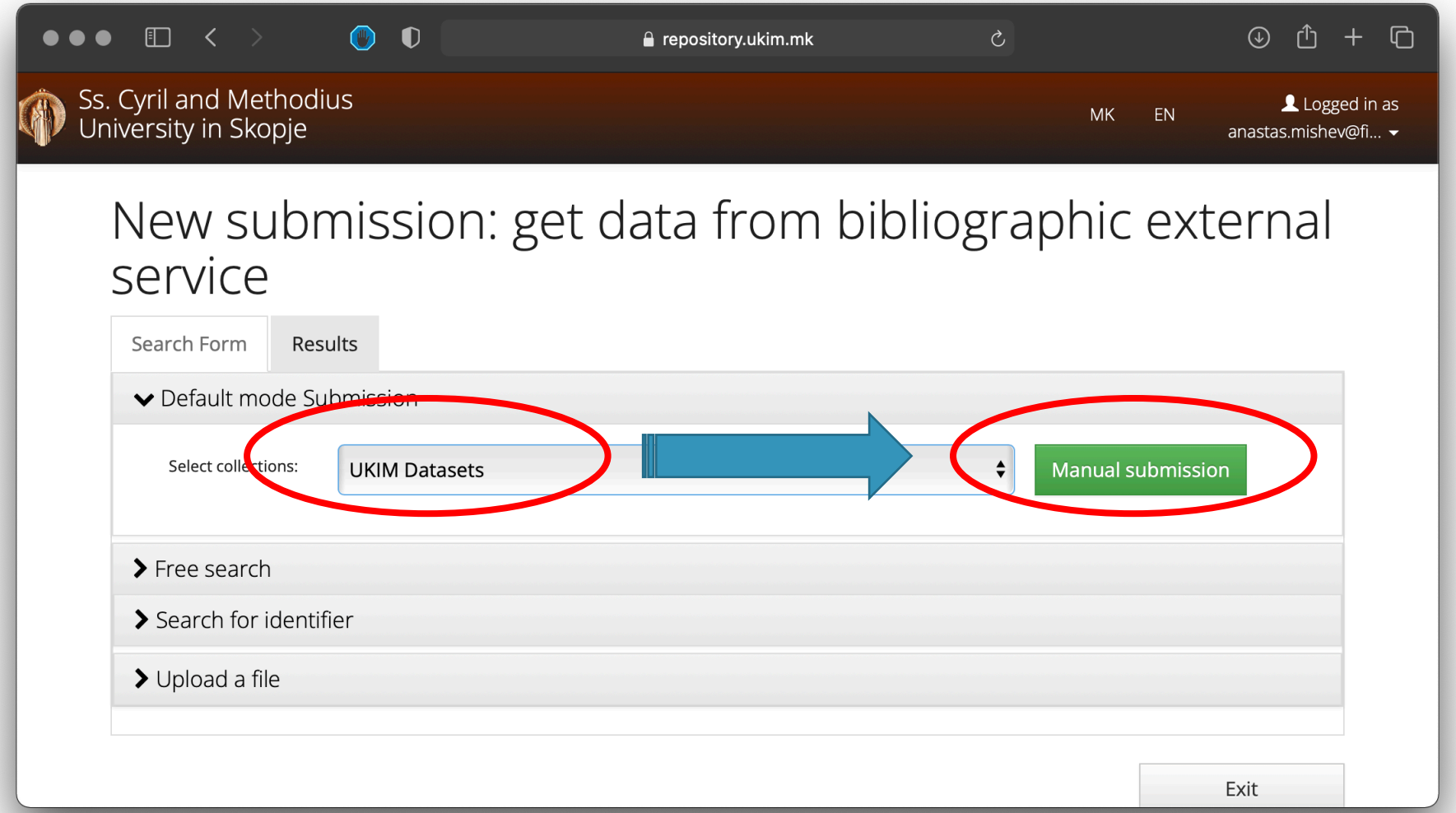
Преглед на тековни записи



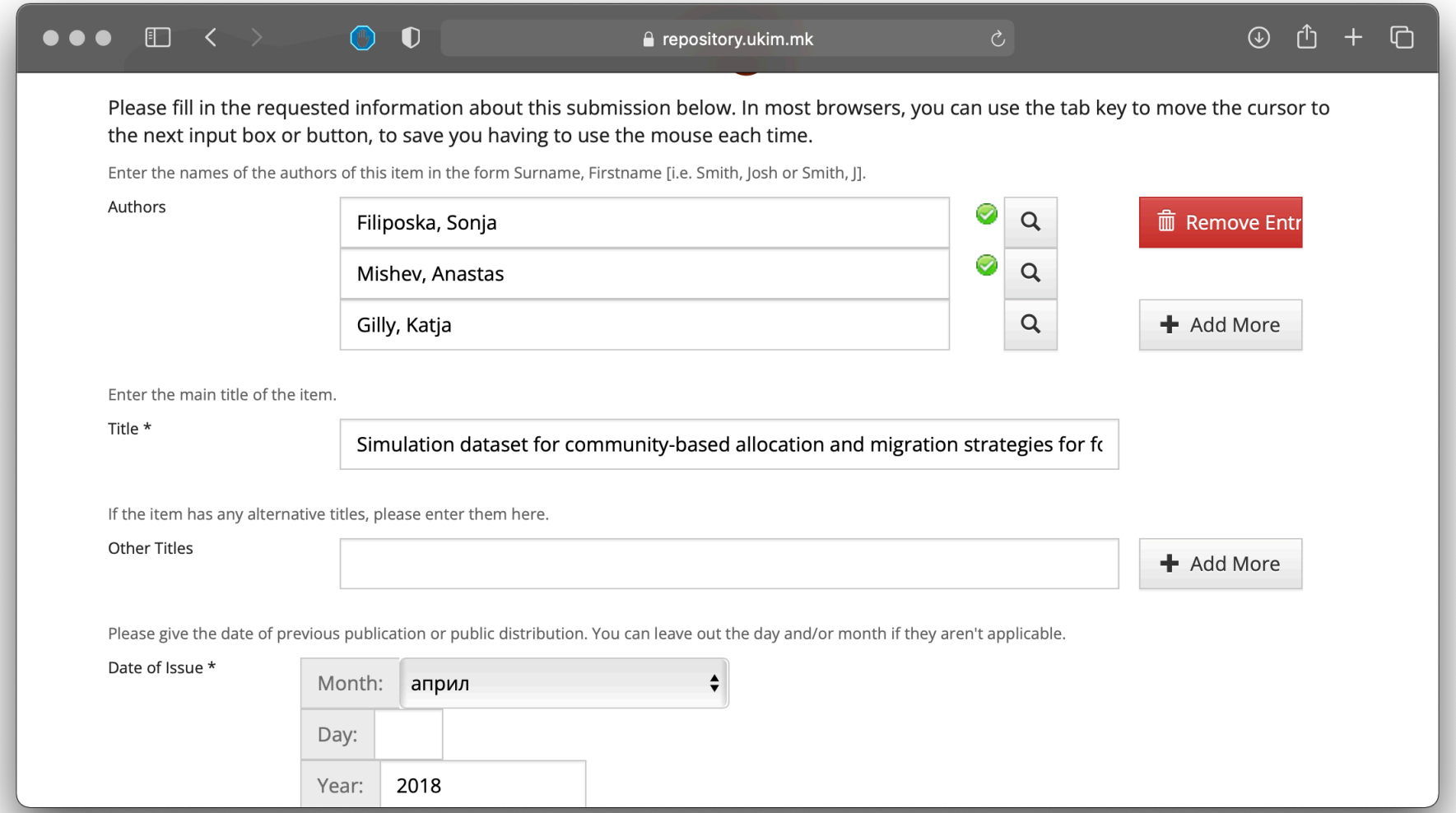
Нов запис


repository.ukim.mk
Ss. Cyril and Methodius University in Skopje
MK
EN
Logged in as anastas.mishev@fi...
REPOSITORY OF UKIM
Repository of UKIM
Search term...
My DSpace: Mishev, Anastas
Researcher profile status: public
Start a New Submission
View Accepted Submissions
Tasks in the Pool
Below are tasks in the task pool that have been assigned to you.
Task
Item
Submitted To
Submitted By

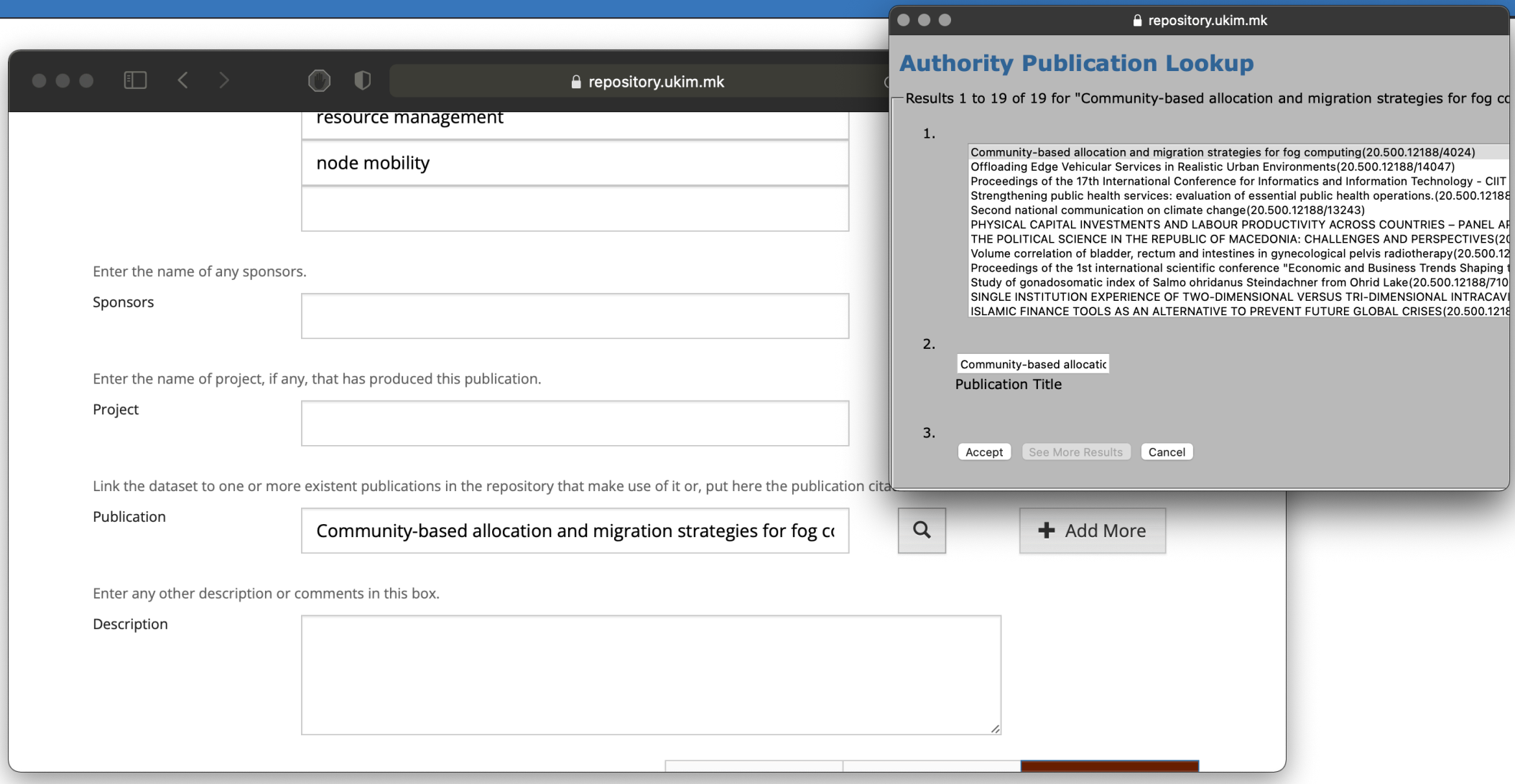
Избор на тип и рачен внес



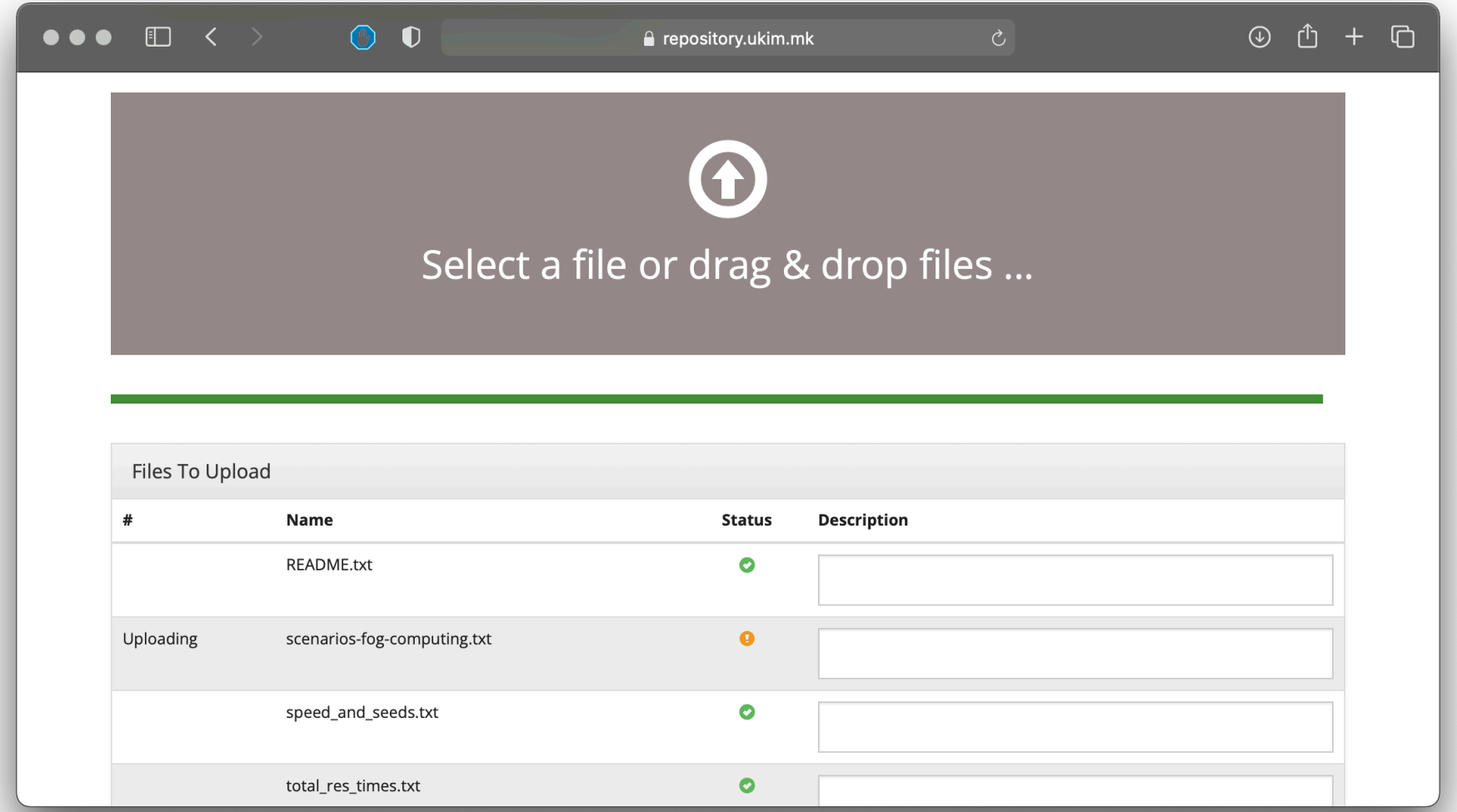
Внесување автори и опис



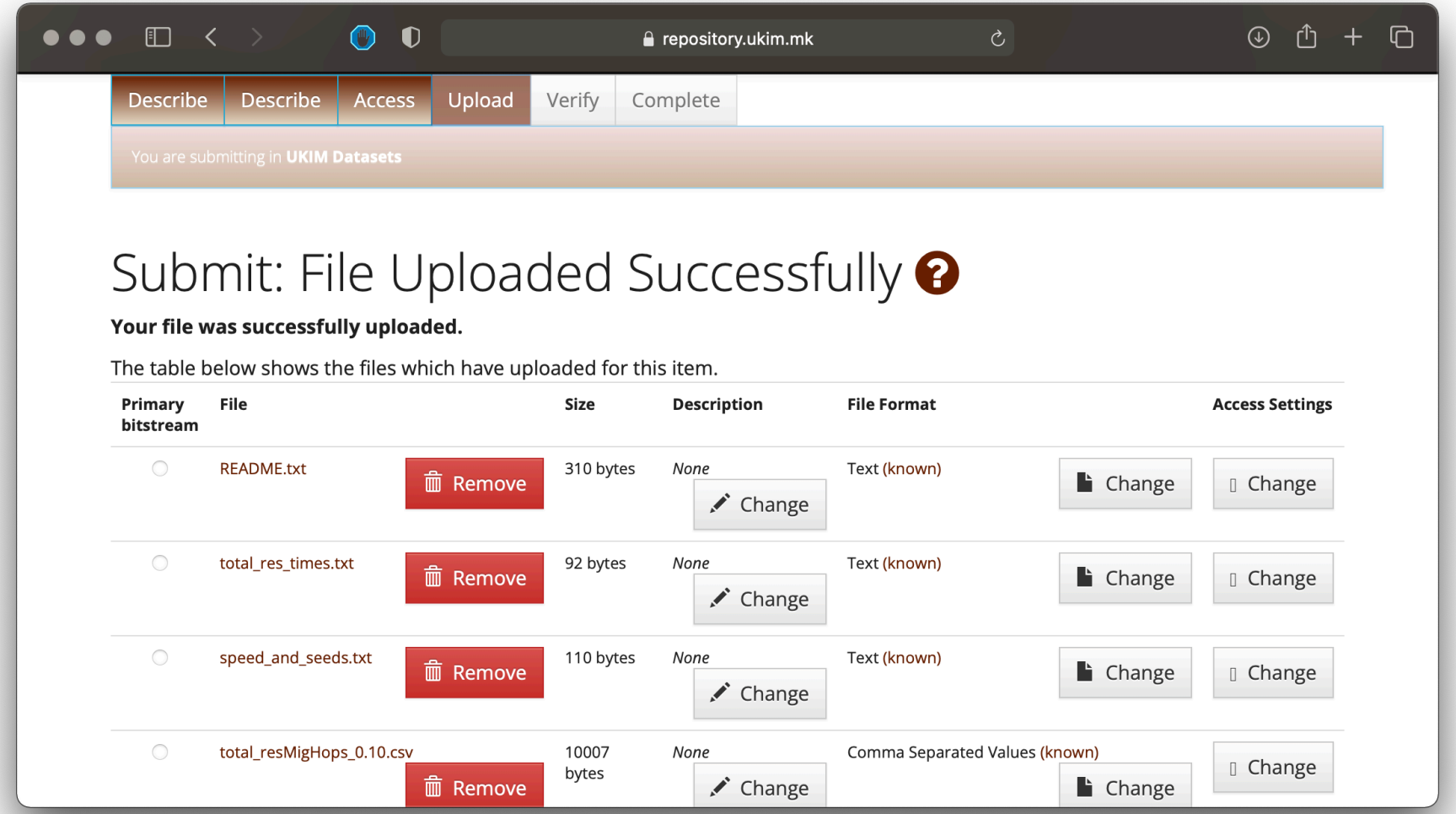
Поврзување со труд



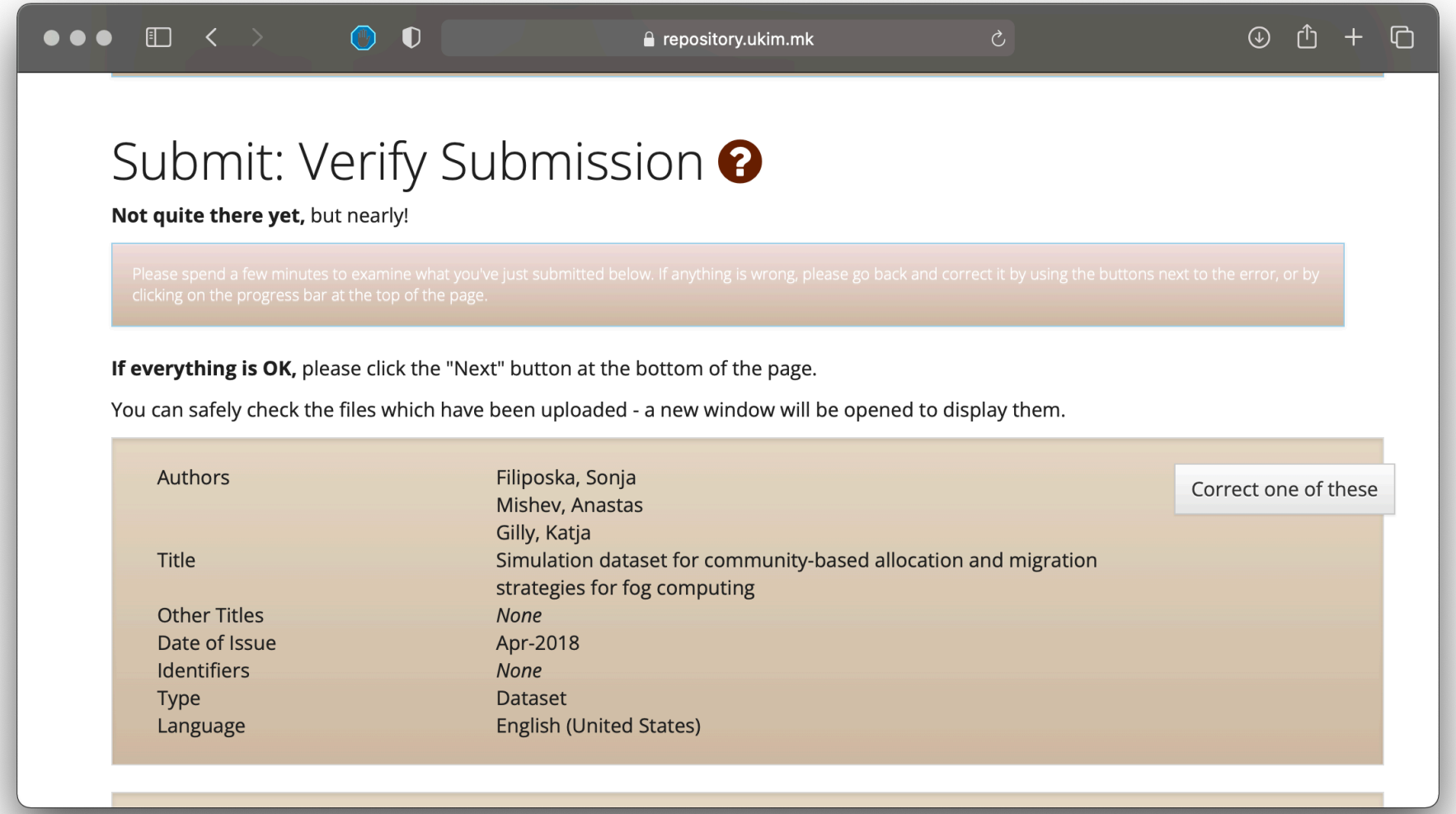
Поставување на датотеки



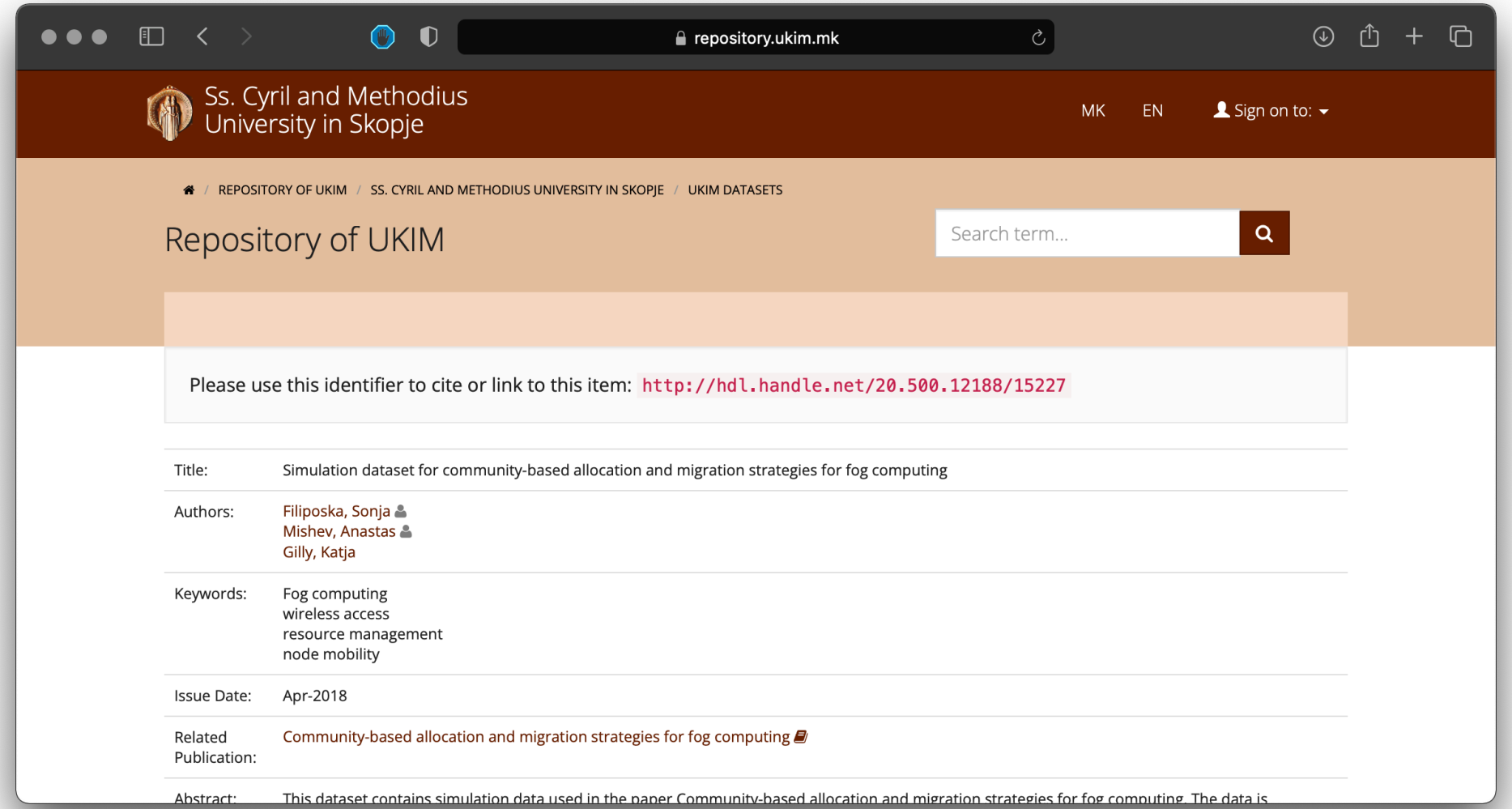
Преглед на датотеки



Верификација

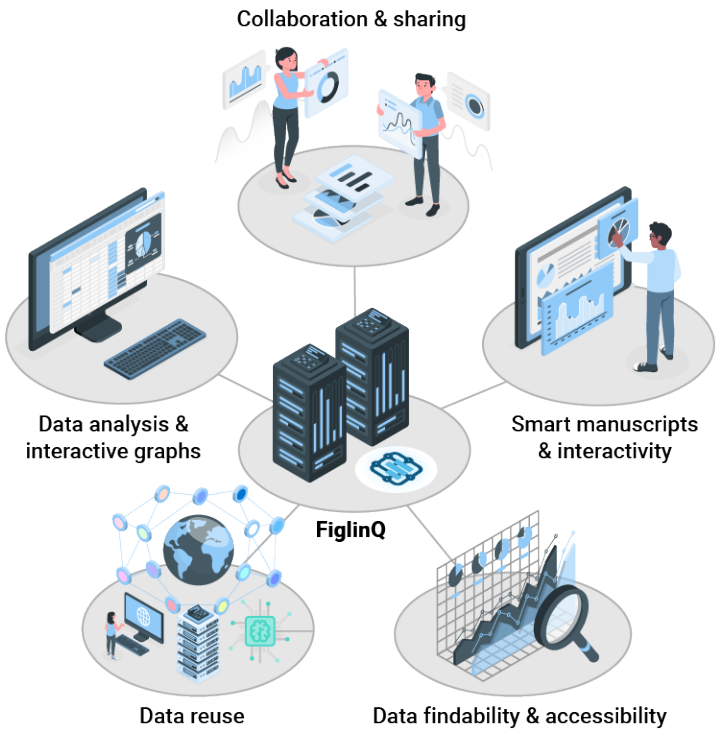


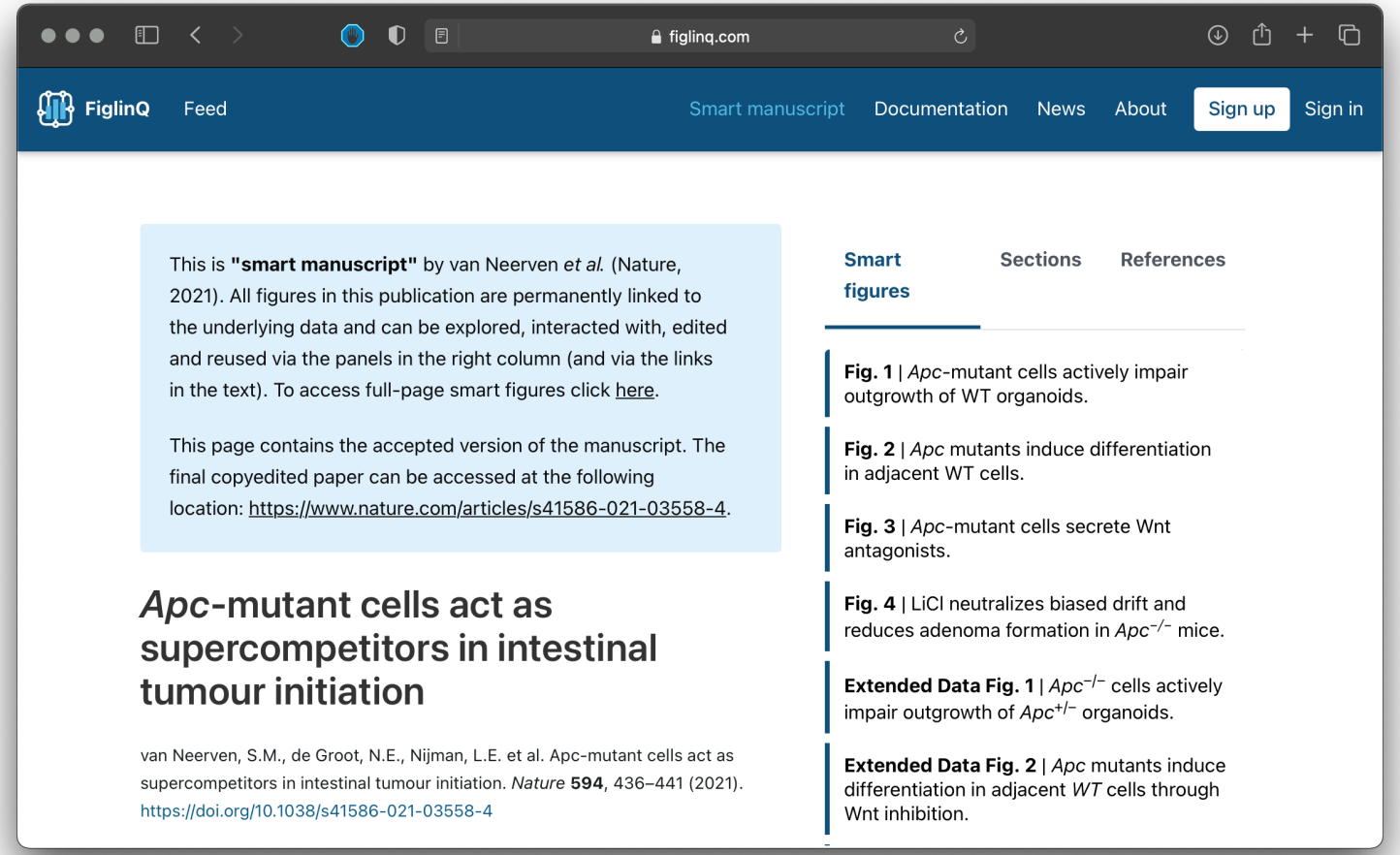
Пристап


repository.ukim.mk
Ss. Cyril and Methodius University in Skopje
MK
EN
Sign on to:
REPOSITORY OF UKIM / SS. CYRIL AND METHODIUS UNIVERSITY IN SKOPJE / UKIM DATASETS
Repository of UKIM
Search term...
Please use this identifier to cite or link to this item: http://hdl.handle.net/20.500.12188/15227
Title: Simulation dataset for community-based allocation and migration strategies for fog computing
Authors: Filiposka, Sonja
Mishev, Anastas
Gilly, Katja
Keywords: Fog computing
wireless access
resource management
node mobility
Issue Date: Apr-2018
Related Publication: Community-based allocation and migration strategies for fog computing
Abstract: This dataset contains simulation data used in the paper Community-based allocation and migration strategies for fog computing. The data is

Иднината на научните публикации

https://figlinq.com/smart

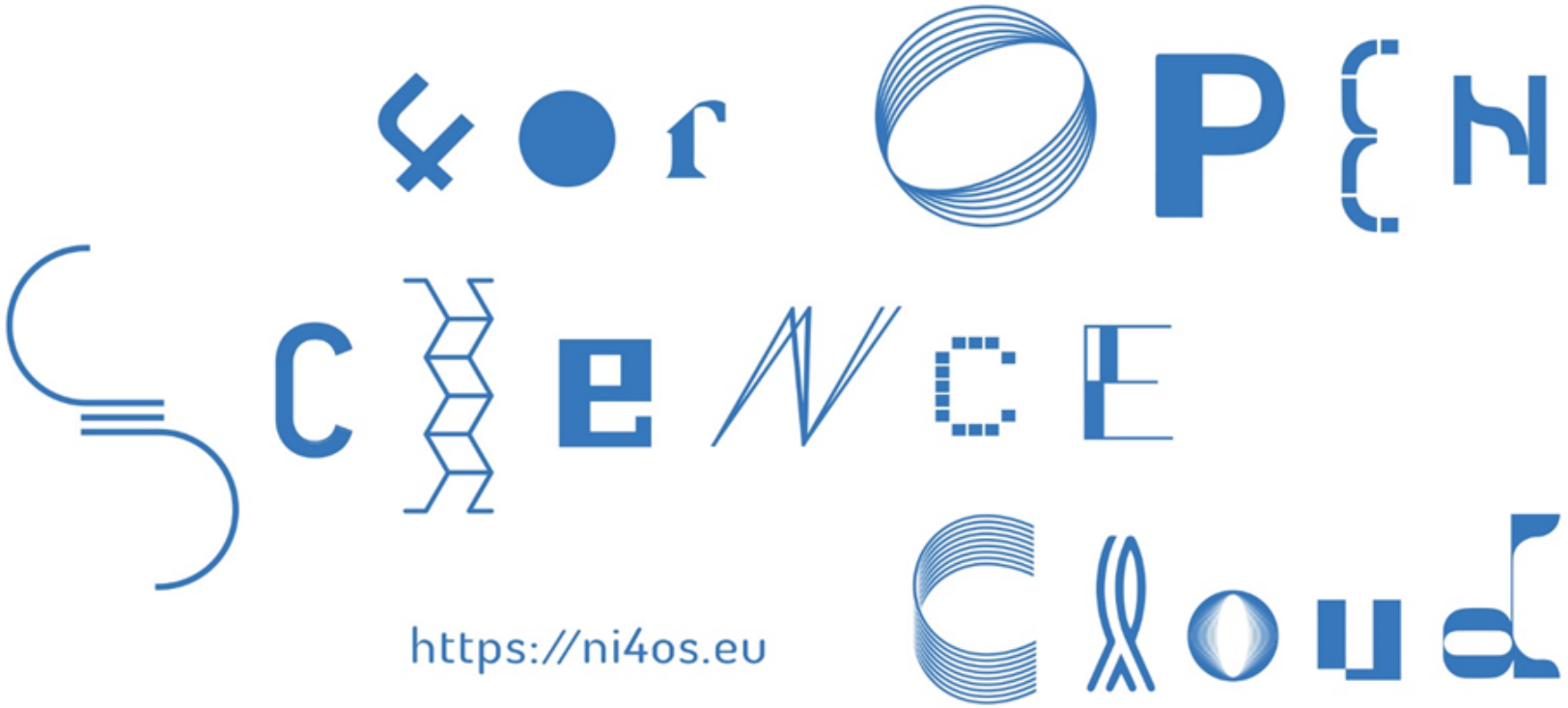




Референци

- https://repository.ukim.mk/handle/20.500.12188/11971
- https://figlinq.com/smart
- https://isps.yale.edu/news/blog/2013/07/the-role-of-data-repositories-in-reproducible-research
- https://digitalguardian.com/blog/what-data-repository
- https://www.astera.com/type/blog/data-repository/
- https://authorservices.taylorandfrancis.com/data-sharing/share-your-data/repositories/
- https://stealthbits.com/blog/what-is-a-data-repository-and-what-is-it-used-for/
- https://journals.asm.org/list-data-repositories
- https://www.nature.com/sdata/policies/repositories
- https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en
- https://like-itn-digitalization.readthedocs.io/en/latest/2_open_science/

Благодарам!



NI4OS_eu NI4OS ni4os.eu