

DATA EXPERTS TRAINING PROGRAMME – AUTUMN 2024

Week 1 (23–26 September 2024)		
Day 1 (23 Sep): OPEN SCIENCE IN GENERAL		
CONTENTS	LECTURERS	DURATION
Fundamentals of Open Science <ul style="list-style-type: none"> • Concept of open science. • Brief history of scholarly communication and open science. • Fundamental principles and values of open science. • Scope of open science (open publications/open access, open data, open software and hardware, open education, citizen science, etc.). 	Miro Pušnik (CTK)	2 hours
Open Science Policies <ul style="list-style-type: none"> • European Research Area (ERA) policies: what ERA is, objectives of the new ERA, three aspects of ERA reform: governance change, Pact for Research and Innovation with 10 shared values, ERA Policy Programme with 20 specific measures, review of measures, Knowledge Exchange Platform (KEP). • Brief overview of Slovenian legislation (ReZrIS30, ZZrID, ZDIJZ, Open Science Regulation). • Brief overview of global policies. • Open Science Action Plan, Measure 6.2. 	Miro Pušnik (CTK)	2 hours
Open Access to Scientific Publications <ul style="list-style-type: none"> • Possible pathways of open access publishing. • Funder requirements (Horizon Europe, Decree on Open Science and funding instruments of ARIS). • Transformative agreements and benefits of open access for Slovenian researchers. • Open access policy of the cOAlition S. 	Miro Pušnik (CTK)	2 hours

Day 2 (24 Sep): FORMAL BASIS FOR THE ROLE OF DATA EXPERTS		
CONTENTS	LECTURERS	DURATION
<p>The Role of Data Experts</p> <p><u>Introductory Information on Working with Research Data</u> Basic information on the fundamental aspects of responsible research data management (explanation of basic concepts, FAIR principles, metadata, persistent identifiers, trusted repositories, exceptions, etc.).</p> <p><u>Characteristics of the Profession/Activities of Data Experts</u> Data experts provide support services to researchers within research projects at specific FTE rates or funded by projects or other sources provided by research organizations. Data stewards work concurrently with multiple researchers, potentially across several institutions, to produce documents and (meta)data. Data librarians may manage COBISS input and repository input depending on the business model. Data consultants advise and support data stewards and librarians as information experts in research data management.</p> <p><u>Defining the Role, Tasks, and Competencies of Data Experts</u> Types of data experts (data stewards, data consultants, data librarians, etc.).</p> <p><u>Task Set for Research Data Management (All Possible Profiles)</u></p> <ul style="list-style-type: none"> • Collaboration with institutional management to ensure compliance with organizational legal requirements. • Information and awareness campaigns for researchers on the importance, methods, and conditions for preparing and sharing FAIR data. • Collaboration with researchers on: <ul style="list-style-type: none"> ○ Preparation of research data management plans (DMPs), ○ Monitoring the data lifecycle, ○ Preparation of data for publication (formats, protocols, provenance, metadata, managing sensitive data, legal frameworks, etc.), ○ Selection and evaluation of repositories, ○ Publication in repositories, 	<p>Bernarda Korez (UKM), Dunja Legat (UKM)</p>	<p>3 hours</p>

- Preparation of reports for funders.
- Bibliographic data processing.
- Preparation of data and documents for evaluating scientific research work (according to funder/organizational needs).

Education and Necessary Skills of Data Experts

Preferably PhD or at least VII/2 for stewards and consultants + experience with research, VII/2 for librarians.

Essential skills include:

- General knowledge of open science,
- Science communication under open science conditions (publication methods – green/gold/diamond/rights retention strategy),
- Knowledge of EU and Slovenian legislation, funders' policies,
- Copyright and licensing laws related to open publications and research data,
- Data protection regulations,
- Requirements of funders for research data (DMP preparation, publication, etc.),
- Preparation and updating of DMPs,
- FAIR principles,
- Data organization (naming conventions, folder structure), data formats (proprietary/open),
- Basics of metadata schemas, vocabularies, ontologies,
- Understanding of repository operations and credibility requirements (persistent identifiers, storage, policies, security, etc.),
- Methods/procedures for data publication (repositories, data papers, metadata, citing data),
- Ability to deliver training on research data management (and open science in general),
- Understanding of ethics and integrity in science,
- Basic knowledge of data science, database work, data mining, machine learning,
- Basic statistics and information visualization (as per research area),
- Basic programming skills (command line, Python, R) and support services (GitHub, etc.) – as per organizational needs,
- Bibliographic processing and COBISS entry competencies (for data librarians).

<p><u>Specificities of Data Experts' Work in Natural Sciences, Engineering, Medicine, Social Sciences, and Humanities</u> Major differences include engineering/natural sciences (use of instruments, large digital data) vs. social sciences/humanities (surveys, analyses, personal data protection); specificity of medicine (data acquisition with instruments, personal data protection).</p> <p><u>Formal Position of Data Experts in Research Organizations</u> Includes salary, career advancement, formal training (varied forms in Slovenia and abroad).</p> <p><u>CONCLUSIONS</u> Define necessary competencies and skills through short discussion as a starting point for business model debates.</p>		
<p>Data Experts in Practice A lecture on the organisation of data experts' work at a European university – an example of good practice.</p>	Antti Rousi (Aalto University, Finland) – online	1 hour
<p>Developing Support Services for Data Experts Open debate on a tailored approach depending on the needs and organizational structure of individual research organizations.</p> <p>How this profession/service will be organised in Slovenia:</p> <ul style="list-style-type: none"> • Individual institutions may hire data stewards based on their specific scientific requirements. • If the demand within an institution is insufficient (less than 1 FTE), two or more institutions could jointly employ a shared data expert. • Multiple institutions could collaborate in a consortium to collectively hire suitable data experts (consortium management may be an issue). • The funding organization or its agency could employ data experts to serve research organisations (time allocation, priorities, and institutional integration may be an issue). <p>Organising support for data experts within the project (support group).</p>	Bernarda Korez (UKM), Dunja Legat (UKM)	2 hours
<p>WORKSHOP: Creating business models for data experts' role for different types of institutions.</p>	moderated by Dunja Legat (UKM)	Not specified

Day 3 (25 Sep): EDUCATION AND INTEGRITY		
CONTENTS	LECTURERS	DURATION
Planning and Delivering Education and Training for Users <ul style="list-style-type: none"> Identifying training needs Tailoring training to meet user needs Collaborating with researchers: communication, networking, etc. Andragogy and didactics 	Iryna Kuchma (EIFL, Lithuania)	3 hours
Ethics and Integrity in Science <ul style="list-style-type: none"> Good scientific practice (research integrity). Reproducibility. Data manipulation and distortion. <i>Cherry picking, HARKing (Hypothesizing After the Results are Known), P-hacking.</i> Quality control (data validation) – rules/recommendations for data quality. 	Urša Opara Krašovec (FE UL)	2 hours
Untrustworthy Publishing Practices Overview of untrustworthy publishing practices and description of other problems in modern science (e.g., scientific paper mills, citation cartels, etc.); incentives for unethical behaviour.	Miro Pušnik (CTK)	1 hour

Day 4 (26 Sep): LEGAL ASPECTS OF OPEN SCIENCE		
CONTENTS	LECTURERS	DURATION
Copyright <ul style="list-style-type: none"> Copyright basics, Open licences, Copyright and scientific publications (open publications), Copyright and research data, Barriers to open science in Slovenian copyright laws, 	Maja Bogataj Jančič (IPI)	4 hours

<ul style="list-style-type: none"> Eligible exemptions from openness. 		
Software Licensing <ul style="list-style-type: none"> (Open) software licences. 	Matija Šuklje	1 hour
Protection of Personal Data (GDPR) <ul style="list-style-type: none"> Lawfulness and ethics. 	Nina Komočar Urbanija (UL)	2 hours

Week 2 (14–17 October 2024)		
Day 5 (14 Oct): RESEARCH DATA LIFE CYCLE 1		
CONTENTS	LECTURERS	DURATION
Definition and Life Cycle of Research Data, FAIR Principles <ul style="list-style-type: none"> Definition of research data (working with data throughout the research lifecycle). Research data lifecycle (planning, acquisition, processing and preparation for publication, publication, reuse). Detailed explanation of the FAIR principles. 	Ana Slavec (FAMNIT UP)	1 hour
Design and Generation of Research Data <ul style="list-style-type: none"> Review of existing data and assessment of potential reuse of other/secondary data (for additional analysis or comparison) Selection of methods, instruments, expected data, measurement parameters, and data quality assessment Creation of metadata (using standardized metadata schemas, vocabularies, and ontologies) Provenance and data description Planning data processing (number of stages, processing methods, software requirements, need for enriching secondary data) Processing of raw data (calibration, standards, normalisation, geolocation, impact on data quality, use of own/foreign algorithms, methods/software, algorithm/program testing...) 	Ana Slavec (FAMNIT UP)	2 hours

<p>Processing and Preparing Research Data for Publication</p> <ul style="list-style-type: none"> • Enrichment/annotation of existing (secondary) data. • Statistical methods and quality assessment (accuracy, precision, repeatability, etc.). • Encryption, anonymisation, pseudonymisation, data protection. • File naming, data organisation, versioning. • Data formats, conversion of proprietary formats to open formats. • Ethics (role of ethics committees, ethical and legal documents) and data quality assurance. 	<p>Irena Vipavc Brvar (ADP, FDV UL), Peter Čerče (ZRS Koper)</p>	<p>3 hours</p>
---	--	----------------

Day 6 (15 Oct): RESEARCH DATA LIFE CYCLE 2		
CONTENTS	LECTURERS	DURATION
<p>Publishing Research Data</p> <ul style="list-style-type: none"> • Publishing data in different types of repositories. • Overview of repositories: <ul style="list-style-type: none"> ○ types of repositories (general, domain-specific, institutional), ○ trusted repositories, ○ conditions (persistent identifier, persistent storage and security, policy, quality assurance, etc.), ○ standardised metadata schemas (repository-specific), ○ duration of data/metadata retention (including any embargoes), ○ publication of comprehensive (rich) metadata and/or data descriptions (ReadMe files) and provenance. • Restrictions on open publishing (personal data, patents, etc.) and access to such data. • Licences for data (CC BY) and metadata (CC0). • Meeting journal requirements for data accessibility in scientific publications (citation methods, access during peer review, publisher requirements for repository submission, etc.). • Data journals and data papers. 	<p>Ana Slavec (FAMNIT UP)</p>	<p>2 hours</p>
<p>Publication of Other Research Results</p> <ul style="list-style-type: none"> • Other types of outputs: software, algorithms, glossaries/thesauri/ontologies. • Persistent identifiers for other research outputs. 	<p>Miro Pušnik (CTK)</p>	<p>1 hour</p>

<ul style="list-style-type: none"> • Software licences. • Licences for other types of objects (CC0 is preferred for glossaries/thesauri). 		
Reuse of Research Data (as a Life Cycle Summary) <ul style="list-style-type: none"> • Finding and reusing data from other researchers (where to obtain it – EOSC). • Persistent identifiers and data citation. 	Brane Leskošek (ELIXIR-SI, IBMI, MF UL)	1 hour
Practical Examples of Working with Research Data <ul style="list-style-type: none"> • One example from natural sciences/engineering, • One example from social sciences/humanities, • One example from medicine. 	CTK, ZRC SAZU, MF UL	2 hours

Day 7 (16 Oct): SPECIFIC DATA MANAGEMENT TOPICS FOR INDIVIDUAL ACADEMIC DISCIPLINES (three concurrent sessions)

Natural Sciences and Engineering

CONTENTS	LECTURERS	DURATION
Data Formats, Metadata, Glossaries, Taxonomies ... <ul style="list-style-type: none"> • Data formats in natural sciences and engineering as a product of instruments: single measurements, measurements dependent on one variable (raster/vector graphs, etc.), dependent on two variables (surface measurements, images, etc.), dependent on three variables (spatial fields). • Standardised metadata schemas in natural sciences and engineering. • Glossaries, ontologies and taxonomies in natural sciences and engineering. 	Uroš Kunaver (CTK)	1 hour

<p>Provenance All the data needed to uniquely define the measurement conditions or reproduce measurements of the same sample under the same conditions.</p> <ul style="list-style-type: none"> The importance of various stages in data processing (raw > normalised/standardised > classified ...) and a detailed description of the protocols/processes/tools/software for transitioning between these stages. Importance of specifying the standards used (in chemistry), calibration protocols/procedures, test/calibration data sets, etc. 	Uroš Kunaver (CTK)	1 hour
Workflows in Natural Sciences and Engineering	Tanja Dreo (NIB), Ana Vučurović (NIB)	1 hour
Spatial Data Aspects	Matija Krivic (GeoZS)	1 hour
Data Stewardship at Dutch Technical Universities	Yan Wang (TU Delft), Bjørn Bartholdy (TU Delft)	1 hour
Panel on Research Data in Natural Sciences and Engineering	moderated by Uroš Kunaver (CTK)	1 hour

Social Sciences and Humanities		
CONTENTS	LECTURERS	DURATION
Data in Social Sciences and Humanities	Andrej Pančur (INZ)	1 hour
<p>Provenance All the information needed to uniquely identify the conditions under which the data were collected and processed.</p>	Andrej Pančur (INZ)	1 hour
<p>Legal and Ethical Aspects Informed consent, protection of personal and sensitive data, anonymisation, data security, protocols for accessing non-anonymised data (safe rooms); copyright and related rights, protection of intellectual property...</p>	Irena Vipavc Brvar (ADP, FDV UL)	1 hour

Standardised Metadata Schemas in Social Sciences and Humanities	To be announced	1 hour
Glossaries, Ontologies and Taxonomies in Social Sciences and Humanities	To be announced	1 hour
Data Publication Preparation for publication, data papers/journals, repositories.	Sonja Bezjak (ADP, FDV UL), Andrej Pančur (INZ)	1 hour

Medicine and Health Sciences		
CONTENTS	LECTURERS	DURATION
Data in (Bio)medicine and Life Sciences <ul style="list-style-type: none"> • Data formats in biomedicine and life sciences – generated by laboratory equipment (digital files). • Persistent identifiers, metadata, vocabularies and ontologies in biomedicine and life sciences. • Metadata standards and schemas in biomedicine and life sciences. • Data traceability and extraction processes in biomedicine and life sciences, including sensitive data. • Archiving, preserving, publishing and sharing biomedical and life sciences data. • Legal and ethical aspects and licensing in biomedicine and life sciences. 	Brane Leskošek (ELIXIR-SI, IBMI, MF UL), Nadja Žlender (ELIXIR-SI, IBMI, MF UL)	6 hours

Day 8 (17 Oct): INFRASTRUCTURE		
CONTENTS	LECTURERS	DURATION
National Open Science Infrastructure and the Regulatory Framework <ul style="list-style-type: none"> • Definition of research infrastructure in the Scientific Research and Innovation Activities Act (ZZrID). • Research Infrastructure in ERA > ERIC > ESFRI and EOSC. • Large international research infrastructure consortia (ERICs), which in Slovenia are often directly linked to open science (CESSDA, CLARIN, DARIAH, Elixir). • A brief presentation of legislation related to open science infrastructure. • A brief presentation of the ARIS criteria for trusted repositories. • Mention of different types of repositories. 	Peter Čerče (ZRS Koper)	0,5 hours

<p>National Open Science Portal</p> <ul style="list-style-type: none"> • History of the National Open Science Portal. • Structure, composition and governance. • FAIR principles; metadata, persistent identifiers, linking to international infrastructures. • Additional features and innovations. • Review of processes to support compliance with open science principles. • Presentation of the Open Science Action Plan for the implementation of Measure 6.2: A6.2.1/3.9: Establish two separate data centres for the long-term storage of research data. 	<p>Milan Ojsteršek (FERI UM)</p>	<p>1 hour</p>
<p>Repositories</p> <ul style="list-style-type: none"> • More detailed presentations of national repositories (ADP, CLARIN.SI, RUL, DKUM, RUP, RUNG, DiRROS, ReVIS). • A brief presentation of examples of high-quality, trustworthy domain-specific data repositories from abroad for selected domains where they do not (yet) exist in Slovenia. • A brief presentation of examples of good quality general data repositories from abroad. • Presentation of infrastructures that are not (yet) repositories in the full sense of the word but are important for open science in Slovenia (ELIXIR-SI, SI-DIH, Sistory, eGeologija ...). 	<p>Janez Brezovnik (FERI UM), Irena Vipavc Brvar (ADP, FDV UL), Brane Leskošek (ELIXIR-SI, IBMI, MF UL), Andrej Pančur (INZ), Milan Ojsteršek (FERI UM)</p>	<p>2,5 hours</p>
<p>Open Access Academic Publishing Platforms and Open Access Science Journals</p> <ul style="list-style-type: none"> • Open Science Action Plan – presentation of Measure 6.2.6: Open access national academic publishing. • Open-access institutional academic publishing platforms (e.g., UL Publishing House journals, ZRC SAZU Publishing House, University Press of the University of Maribor ...). • The importance of open licences. 	<p>Jonatan Vinkler (UP)</p>	<p>0,5 hours</p>
<p>Other Digital Services and Resources Needed or Generated During Research in Accordance with the Principles of Open Science</p> <ul style="list-style-type: none"> • Open access content platforms relevant to science. • Information and educational platforms focused on open science content (OPSI, e-torba, dLib, etc.). • SLING. 	<p>Peter Čerče (ZRS Koper)</p>	<p>0,5 hours</p>
<p>European Open Science Cloud (EOSC) and Similar Platforms and Data Spaces</p> <ul style="list-style-type: none"> • The concept, meaning and vision of EOSC and its links to the global network. 	<p>Milan Ojsteršek (FERI UM), Marko Drobnjak</p>	<p>1 hour</p>

<ul style="list-style-type: none"> • Operation of the EOSC Association. • Tripartite management. • Working groups. • Slovenia's participation in EOSC. • EOSC portal <ul style="list-style-type: none"> ○ purpose and development, ○ portal elements, ○ possible uses. • Similar platforms and data spaces <ul style="list-style-type: none"> ○ Common European Data Spaces, ○ Global Open Science Cloud (gosccloud.net). 	(ARNES), Irena Vipavc Brvar (ADP, FDV UL)	
--	--	--

Week 3 (18–21 Nov 2024)		
Day 9 (18 Nov): RESEARCH DATA MANAGEMENT PLANS		
CONTENTS	LECTURERS	DURATION
<p>Developing a Research Data Management Plan (DMP)</p> <ul style="list-style-type: none"> • Funder requirements (within the context of the overall research project workflow). • Development of internal policies (rules on managing survey data, rules on quality assurance/data integrity). • Components of research data management plans (templates) – variations across DMPs for different calls/funders (Horizon Europe, Marie Curie, ERC, ARIS, etc.). • Tools for creating DMPs (various templates selection). • Filling out and reviewing a model DMP (based on the translated model available on the DiRROS Data website). • Estimating the costs of appropriate research data management. 	Miro Pušnik (CTK), Peter Čerče (ZRS Koper), Brane Leskošek (ELIXIR-SI, IBMI, MF UL)	5 hours
<p>Preparation of Reports to Funders and Evaluation of Research Data Management Plans</p>	Urša Opara Krašovec (FE UL)	1 hour

Day 10 (19 Nov): DATA ANALYSIS AND MANAGEMENT TOOLS 1		
CONTENTS	LECTURERS	DURATION
Working with Databases (Bibliographic, Factual, Data, Repositories)	Jerneja Grašič (UKM), Dunja Legat (UKM)	1 hour
Introduction to the SPSS Software Tool	Irena Vipavc Brvar (ADP, FDV UL)	1 hour
Introduction to the Phyton and R Software Tools	Maja Zagorščak (NIB)	1 hour
GIS Tools	Simon Koblar (UIRS)	1 hour
Text Encoding initiative, Oxygen and Other Tools	Andrej Pančur (INZ)	1 hour
Selected Tools for Natural Sciences and Engineering	To be announced	1 hour

Day 11 (20 Nov): DATA ANALYSIS AND MANAGEMENT TOOLS 2		
CONTENTS	LECTURERS	DURATION
Laboratory Data Management – Similarities and Differences Between LIMS and ELN Systems and Solutions on the Market	Jana Erjavec (Biosistemika)	1 hour
Tools for Managing Project Data – pISA-tree	Marko Petek (NIB)	1 hour
Tools for Managing Workflows	Marko Vidak (ELIXIR-SI, IBMI, MF UL), Brane Leskošek (ELIXIR-SI, IBMI, MF UL), Maja Zagorščak (NIB)	1 hour
Versioning and Using Git, GitHub/GitLab	Katja Meden (INZ)	1 hour
File/Data Format Conversion Tools	Uroš Kunaver (CTK), Miha Peče (ZRC SAZU)	1 hour
Other Tools – A Brief Overview of Various Other Tools	Uroš Kunaver (CTK)	1 hour

Day 12 (21 Nov): PRACTICAL WORKSHOPS IN GROUPS/ACADEMIC DISCIPLINES

CONTENTS	LECTURERS	DURATION
<p>Developing a Research Data Management Plan Data lifecycle review, research design, repository confidence level assessment, repository selection, metadata selection, data description, data management plan updates during the project, processing and preparation for publication.</p> <p>The workshop will be delivered in three separate sessions for each of the academic disciplines:</p> <ul style="list-style-type: none"> • social sciences and humanities, • natural sciences and engineering, • biomedical sciences. 	ZRC SAZU, ADP, CTK, MF UL, NIB	6 hours