



Universiteit
Leiden

Data Science in

Fieldlabs

16.09.2021

Fieldlab Requirements

The COVID-19 pandemic has caused a revolution in on-line activity. More than ever before digital data is generated and growing exponentially. There is high expectation that such data and the analytics enabled through it with AI, are enabling solutions. At the same time, there is increasing concern that the quality of such solutions may be limited due to the incompleteness of data; that some populations are not included in it and do not want to be included and therefore solutions based on data is based on social biases; and that the use of such data is not necessarily in the public interest or harms particular groups of people.

The use of data in a real life situation is therefore complex and requires understanding of the particular situation of the problem, how the problem is understood by stakeholders, insight in the ethical issues that could emerge, and understanding of the limitations and possibilities of the data-powered solutions.

This course will provide you tools on how to approach a real-life problem and will give you the competence to plan a real-life project.

In this course you will also gain experience, and you are allowed to make errors and to discover the pitfalls of working on such projects in a real-life situation.

In real-life, you are likely to work in an interdisciplinary team with members of different specialisations; this course practices team-work within the context of a situation with stakeholders who have identified problems that need to be solved.

The legal requirements and norms around data-processing have changed rapidly in the last few years. In Europe the General Data Protection Regulation (GDPR) and the principles of Findable, Accessible, Interoperable and Re-usable (FAIR) -Data provide a new framework for the practices of data-analytics.

Data Scientists are now, more than ever before, required to be able to understand their solutions in the context of such new norms as well as be able to advise and reflect on data curation and management that is defined in terms of its purpose and the life-cycle of the data.

This course provides students with a practical experience in which competences are gained to manage a data science problem within a team and to come up with realistic proposals that are bound by the relevant regulations and guidelines.

Programme Fieldlabs

Fieldlab 1		11
Name	PHT	
Organisation	LUMC	
Mentor	Erik Flikkenschild	
Contact details	Erik Flikkenschild <e.flikkenschild@lumc.nl>	
Description of the problem	Develop a basic architecture with requirements and specifications for a PHT in LUMC	
Skills to be acquired and required	modeling	
Degree of challenge	size/volume of the challenge	
Number of Team members	TBD	

Fieldlab 2		1
Name	PHT	
Organisation	LUMC	
Mentor	Olzhas Aldabergenov; Erik Flikkenschild	
Contact details	olzhas.algabergenov@gmail.com e.flikkenschild@lumc.nl	
Description of the problem	Deployment of International localised CEDAR software for FAIR Data Production	
Skills to be acquired and required	modeling	
Degree of challenge	size/volume of the challenge	
Number of Team members	TBD	

Fieldlab 3		2
Name	FAIR repositories of vaccine - research data	
Organisation	LUMC	
Mentor	Aliya Aktau	
Contact details	Aliya Aktau <aleka.aktau@gmail.com>	
Description of the problem	Develop metadata templates for a clinical research on a new vaccine developed in LUMC	
Skills to be acquired and required	vocabulary creation in human and machine-readable format	
Degree of challenge	precision and detail and interdisciplinary collaboration	
Number of Team members	TBD	

Fieldlab 4	3
Name	Development of algorithmic queries within health facilities and visualisation
Organisation	VODAN-A
Mentor	Mariam Basajja
Contact details	mariam mutoni basajja <mariam.basajja@gmail.com>
Description of the problem	Development of analytics and visualisation for use in health facilities based on wishes in health facilities
Skills to be acquired and required	SparQL
Degree of challenge	client-responsive analytics and visualisation
Number of Team members	TBD

Fieldlab 5	4
Name	Development of federated algorithmic queries for aggregate analytics
Organisation	VODAN-A
Mentor	Ruduan Plug
Contact details	Ruduan Plug <ruudplug@gmail.com>
Description of the problem	Distributed computation across facilities for data analytics with deidentified data
Skills to be acquired and required	statistical AI
Degree of challenge	statistics
Number of Team members	TBD

Fieldlab 6	5
Name	Permission architecture for federated algorithmic queries and deidentification
Organisation	VODAN-A
Mentor	Putu
Contact details	Dhr. Putu Hadi Purnama Jati (putu.hadi.purnama.jati) <putu.hadi.purnama.jati@umail.leidenuniv.nl>
Description of the problem	authorisation and security automated contracts
Skills to be acquired and required	automated contracts and authorisation architecture
Degree of challenge	innovation
Number of Team members	TBD

Fieldlab 7		6
Name	Data interoperability on COVID-related health programmes	
Organisation	Philips Foundation	
Mentor	Yi Lin	
Contact details	Lin, Y. (Yi) <y.lin@liacs.leidenuniv.nl>	
Description of the problem	BI model and underlying template production	
Skills to be acquired and required	visualisation and FAIR Data Science	
Degree of challenge	modeling	
Number of Team members	TBD	

Fieldlab 8		7
Name	tbd	
Organisation		
Mentor		
Contact details		
Description of the problem		
Skills to be acquired and required		
Degree of challenge		
Number of Team members		

2. Course management

Hey, let me introduce myself as the one responsible for the course!

Prof Dr Mirjam van Reisen

<https://www.universiteitleiden.nl/en/staffmembers/mirjam-van-reisen#tab-1>

<https://www.lumc.nl/org/lu-cid/medewerkers/mirjamvanreisen>

Prof. Dr. Mirjam van Reisen holds the chair FAIR Data Science at the Leiden University Medical Centre. She previously held the chair Computing for Society at the Leiden Institute for Advanced Computer Science. Mirjam van Reisen is also Prof International Relations, Innovation and Care at Tilburg University Faculty of Humanities and Digital Sciences, Department of Culture Studies. From 2012 - 2021 Van Reisen was a member of the Dutch Government Advisory Council on International Relations and Chair of the Committee on Development Cooperation. She chairs various research programmes, among others the advisory committees of the Scientific Centre of the Dutch Ministry of Justice. Van Reisen is the International Coordinator of the Virus Outbreak Data Network (VODAN) - Africa and Asia. She is project leader of numerous programmes funded by NUFFIC, the European Union, NWO and ZonMW funders. She is coordinator of the international research network Globalisation, Accessibility, Innovation and Care (GAIC). She teaches Data Science Field Labs, Regulatory Governance for Data Science and Globalisation and Migration. Van Reisen is Founding Director of the Europe External Policy Advisors based on Brussels. She was also a founding board member of the Philips Foundation. She received the Golden Image Award from President Ellen Johnson-Sirleaf of Liberia.

Mirjam van Reisen:

Email: mirjamvanreisen@gmail.com

WhatsApp: +32479468690

Teaching Assistants

Mariam Basajja

mariam mutoni basajja <mariam.basajja@gmail.com>

Yi Lin

Lin, Y. (Yi) <y.lin.2@umail.leidenuniv.nl>

Ruduan Plug

Ruduan Plug <ruudplug@gmail.com>

2. Lecturers

Prof. Dr. Mirjam van Reisen

Prof. Dr. Mirjam van Reisen holds the chair 'FAIR Data Science' at the Leiden University Medical Centre. She is also Prof 'International Relations, Innovation and Care' at the Tilburg University at the Faculty of Humanities and Digital Sciences. She was previously Professor 'Computing for Society' at the Leiden Institute for Advanced Computer Science. She is the international coordinator of the Virus Outbreak Data Network (VODAN)-Africa. She leads several international research programmes around innovations on FAIR-Data. Mirjam van Reisen is coordinator of the [Research Network Globalisation, Accessibility, Innovation and Care \(GAIC\)](#) a global research network on the interdisciplinary issues of data and computation, medicine and health and social aspects of care. Van Reisen's publications are listed here: <https://mirjamvanreisen.wordpress.com/>

Prof. Dr. Mark Musen

Prof. Dr. Mark Musen is a Professor of Biomedical Informatics at Stanford University, where he is Director of the Stanford Center for Biomedical Informatics Research. Prof. Musen conducts research related to open science, data stewardship, intelligent systems, and biomedical decision support. Prof. Musen chaired the Health Informatics and Modeling Topic Advisory Group for the World Health Organization's revision of the International Classification of Diseases (ICD-11) and he currently directs the WHO Collaborating Center for Classification, Terminology, and Standards at Stanford University.

Prof. Ronald de Jong

Prof. Ronald de Jong is a Distinguished Professor of Practice at Tilburg University with the School of Economics and Management (TiSEM). There he aims to the gap between theory and practice, helping students having a positive impact on the world while ensuring the research we conduct has the right impact on society and applied science. Prof. De Jong has been the Chief Human Resources Officer and the Executive Vice President of Philips. In addition, Prof. De Jong has founded and is a chairman for the Philips Foundation, which uses innovation and talent to reduce healthcare inequality by providing access to quality healthcare for disadvantaged communities.

Dr. Katy Wolstencroft

Dr. Katy Wolstencroft is an assistant professor at Leiden University, where she leads the Semantic Systems Bioinformatics group at the Leiden Institute of Advanced Computer Science (LIACS) and teaches courses in Biomedical Informatics and Computational Molecular Biology. Dr. Wolstencroft's research group focusses on automated data extraction, FAIR data management and semantically integrating biomedical data. Dr. Wolstencroft has previously held a position as a research fellow at the University of Manchester.

Prof. Dr. Francisca Onaolapo Oladipo

Prof. Dr. Francisca Oladipo is a Professor of Computer Science and pioneer Director, Quality Assurance, Federal University Lokoja (FUL), Nigeria. She was a member of the University's Governing Council – the first female elected to the Council, and the first female Head of the Department of Computer Science at the University. She is a Consultant at the Tilburg School of Humanities and Digital Sciences of Tilburg University, Netherlands, and had served previously at Kampala International University (KIU), Uganda as the Director of Research, Innovations, and International relations.

Aliya Aktau

Aliya Aktau is a PhD candidate at Leiden University and also an alumni of the Leiden Institute of Advanced Computer Science (LIACS). Currently she works as a lecturer at the Suleyman Demirel University at the Faculty of Engineering and Natural Sciences. Aktau is a technical coordinator at VODAN Africa & Asia working towards the implementation and deployment of FAIR Data Principles in Africa and Asia.

Mariam Basajja

Mariam Basajja is a PhD candidate at Leiden University and has a 1st class honours degree in Computational Intelligence from the University of Nairobi. Basajja's main research interest are on artificial intelligence and bringing FAIR data principles to practice. Basajja is an ambassador for Uganda at the GO FAIR initiative and is a technical coordinator at VODAN Africa & Asia, where she works on the development of FAIR data standards and stewardship.

Putu Hadi Purnama Jati

Putu is a master graduate from the Leiden Institute of Advanced Computer Science (LIACS) and currently a researcher in VODAN-Africa. He is also working at the National Bureau of Statistics in Indonesia. In VODAN-Africa he is responsible for the interviews conducted with all resource persons and stakeholders in the project.

Erik Flikkenschild

Senior IT advisor Research (LUMC), working on FAIR implementation (quartermaster), Research workspace (VRE) (national reference architecture,) secure data links with personal data (chairman NFU/Surf working group SIG secure data links with personal data).

3. Guest lectures

During this course, guest lectures will be provided by experts in various areas, related to digital data management:

- Mustafa Kediolgu: Project management
- Prof Mark Musen: Bio-informatics and the CEDAR Platform
- Prof Ronald de Jong: Steering projects in practice
- Prof Francisca Oladipo: The Virus Outbreak Data Network programme
- Ruduan Plug: statistical AI of federated data
- Mariam Basajja and Aliya Aktau: creating machine-readable ontologies and SparQL queries with deidentified data and unique identifiers
- Putu Hadi Purnama Jati: Access and control and permission under GDPR
- Dr Katy Wolstencroft: Data management solutions
- Yi Lin: Data visualization and analytics
- Erik Flikkenschild: FAIR implementation, Personal Health Train and data security

4. FAIR Data Science

- “FAIR Data Science” is a collaboration with LUMC, [Philips](#) and [Kampala International University \(KIU\)](#). [Findable, Accessible, Interoperable and Reusable \(FAIR\) Data](#) is the next generation of ethical data stewardship for sciences and services.
- FAIR-data provides a vision of building the Internet of FAIR Data and Services based on machine-readable federated data.
- The network is established in nine African countries to create FAIR interoperable patient data held in residence.
- The machine-readable data can be visited over the internet with permission being linked to data governance in place where data is produced. FAIR-data is compliant with the EU GDPR.
- FAIR also stands for federated AI-Ready. Global ethical data strategies provide a basis for the creation of diverse quality data, which is a key concern for the development of relevant AI.

5. Course Objectives

General Objectives

1. Learning: To have knowledge of the GDPR and FAIR requirements and their application for data science architectures on data acquisition and - analytics and processes for data science and learn about different applications
2. Competences: To evaluate a data science problem and to assess feasible approaches towards a solution within the boundaries of minimum criteria for quality
3. Skills: To exercise new techniques relevant to FAIR Data curation and acquire different toolsets for planning a data-science project and to know how to use these, and to practice skills of presentation, innovation research, interviews and writing.

Specific Objectives

- Studying of relevant literature and documentation on
 - o GDPR-requirements for data-processing
 - o FAIR-guidelines
- Learn how to realistically plan a project that is limited by time and resources, use project planning tools and practice team-work, using tools to optimize interdisciplinary team-work
- Acquisition of practical skills related to tools that relate to data curation according to FAIR- and GDPR-based processes, such as CEDAR template development, SparQL, deidentification, unique identifiers and facets of FAIR-based data curation
- Insight in the relevance of data curation for quality data analytical processes and the advantages and limitations of methods based on machine-readable and semantic web based approaches
- Have increased awareness of the real-life settings of data science challenges and the relevance of context and stakeholders for the development of solutions

Planning

Week	Date	Topic	Objective	Lecture Preparatory Reading	Class	Learning input	Assignment	Grading	Individual/Group work
1	16/09/2021	16:15-18:00	Introduction to Data Science in Practice	Students will have an understanding of the course objectives; course structure; requirements and assignments - and the learning objectives, including competences and inventory on skills for FieldLabs	Mirjam van Reisen, Mariam, Ruduan	Article: Reisen, M., Oladipo, F., Stokmans, M., Mpezamihgo, M., Folorunso, S., & Schultes, E. et al. (2021). Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. <i>Advanced Genetics</i> , 2(2). doi: 10.1002/ggn2.10050	Working session. Division in groups and preparation of a powerpoint. Presentation in plenary	Information and planning	Send list of issues that are not clear and reflections from the article reading in BrightSpace - Assignment 1
2	23/09/2021	16:15-18:00	Evolution in Data Science	Students understand how data science has evolved from a statistical field to an AI domain to a data curation domain, particularly in the field of health	Mirjam van Reisen, Ruduan Plug	CRISP-DM : https://the-modeling-agency.com/crisp-dm.pdf & Plug, R., Liang Y., Aktau, A., Basajja, M., Oladipo, F. O., Van Reisen, M. (2021) Terminology on a FAIR-framework for the Virus Outbreak Data Network Africa. <i>Data Intelligence. Special Issue. Forthcoming.</i>	Class debate on the two approaches	Learning	Short paper on the essential differences between CRISP-DM and FAIR-based data Science
3	30/09/2021	16:15-18:00	Presentation of Field Labs	Students receive all information on field labs and practical tools on how to organise the process of implementation of the Field labs with input on: (i) how to plan a time-bound project (ii) how to organise in a team and (iii) how to create a relevant process for an interdisciplinary data science challenge	Mustafa Kedioglu	Terminology paper Ruduan; Fieldlabs propositions	Guest lecture and preparation of FieldLab Teams	Competences	Select FieldLab of interest
4	07/10/2021	16:15-18:00	SOLID and Data Standards - A FAIR- based approach to health data - BENEFIT - programme	Considering the evolution in Data Curation for Data Science: automated data extraction, FAIR data management and semantically integrating biomedical data and enhancing ability to implement a stakeholders interview in compliance with ethical procedures	Mingyue; Katy Wolstencroft and Putu Hadi Purnama Jati	https://www.nature.com/articles/sdata201618 and https://guides.lib.vt.edu/researchmethods/interviews	Guest lecture and discussion of practical examples and outcomes - including key references to GDPR for considerations of access and permission assessments and roles and responsibilities in data management	Learning and skills	Create a 3-pager to introduce the Fieldlab problem to interviewees and to develop a interview topic list as a tool to interview the stakeholders and understand their expectations and needs. The document should also include a list of stakeholders to interview, contact details and a timeplan for execution of the interviews.
				5%	Group work				

- 5 14/10/2021 16:15-18:00 Innovation in company settings The social responsibility of companies and organisations is considered in light of assessing directions for innovation with a case study of the Philips company in the context of health-services Ronald de Jong, Prof in Practice De Jong, R., Vermeulen, F., The Strategic Transformation of Royal Philips, London Business School. June 2021 Guest lecture and discussion on challenges in company transformation towards Global Development Goals - during this lecture we will also introduce the concept of 'stakeholders' and introduce 'FieldLab mentors'. We will present the next assignment on exploring expectations of Fieldlab mentors and other stakeholders. We will discuss how to practically prepare for interviews and what tools to use (consent, interview topic list, ethical considerations). Competences Create a 2-pager by the Fieldlab Team on the core of the Field-lab problem; the members of the team, the skills available in the team and come up with a Team name. The paper should also set out how you are planning the work in the team, using the tools provided in the Guest Lecture of Mustafa Kedioglu 5% Group work
- 6 21/10/2021 16:15-18:00 Production of metadata and metadata annotation with FAIR-specifications with emphasis on interoperability through machine-readability of metadata and data re-use To identify practical tools and software to produce FAIR-data in the field of health and biomedicine. Mark Musen with assistance of Ruduan Plug Read: FAIR Data Generation of Demonstration of the CEDAR Metadata architecture and toolbox skills Interview-reports 5% Group work
- 7 28/10/2021 16:15-18:00 The Virus Outbreak Data Network (VODAN)-Africa Students will be informed about the challenges of an innovation and what may contribute to the success of the innovation, including SMART goals Francisca Oladipo and Aliya Aktau SMART-planning tools Feedback of interviews; lecture and introduction of SMART Planning tools and short five minute presentations of the FieldLabs skills Prepare a GANNT chart and planning tools with SMART objectives for the execution of the project and short presentation of the project 5% Group work
- 8 04/11/2021 16:15-18:00 Skills development 1: Interoperability through FAIR-Template creation in CEDAR, SparQL, deidentification and unique identifiers Students will acquire the competences Mariam Basajja, Aliya Aktau, Yi Lin, Ruduan Plug provided by trainers Group presentations and peer review skills Short presentation on the Field Lab ambitions and planning 5% Group work
- 9 11/11/2021 16:15-18:00 Skills development 2: Interoperability through FAIR-Template creation in CEDAR, SparQL, deidentification and unique identifiers Students will acquire the competences Mariam Basajja, Aliya Aktau, Yi Lin, Ruduan Plug provided by trainers Work in computerlab in groups skills
- 10 18/11/2021 16:15-18:00 Personal Health Train LUMC, accessibility, permissions and interoperability issues Reflecting on accessibility and security issues based on a case study of FAIR-Innovation in Leiden University Medical Centre to generate ideas on possible solutions for Fieldlab problems Guest lecture by Erik Flikkenschild Purnama Jati, P.H., Lin, Y., Cohyono, D. B., Nodehi, S., & van Reisen, M. (2021). FAIR and Open Data differences. Data Intelligence. Special Issue. Forthcoming. and Purnama Jati, P.H., van Reisen, M., Flikkenschild, E., Oladipo, F. O., Meerman, B., Plug, R., & Nodehi, S. (2021). Data Access, Control, and Privacy Protection on VODAN Africa Architecture. Data Intelligence. Special Issue. Forthcoming. Discussion on permission and access challenges competences

- 11 25/11/2021 16:15-18:00 Visualisation and BI approaches Discovering topics to consider for the data analytics of data generated including issues of data ownership and introduction on presentation and final paper Yi Lin Purnama Jati, P.H., Lin, Y., Cohyono, D. B., Nodehi, S., & van Reisen, M. (2021). FAIR and Open Data differences. Data Intelligence. Special Issue. Forthcoming. Demonstration of BI solutions and considerations followed by discussion in groups competences
- 12 02/12/2021
- 13 09/12/2021 16:15-18:00 Review & revision and preparation for final paper Students have clarity on requirements of final assignments of the course, criteria for passing and timelines Mirjam van Reisen <http://www.jsu.edu/depart/geography/mhill/research/researchf.html>
Presentation and discussion on instructions for individual final paper, presentation of outline and requirements, discussion of examples.
Instructions and preparation of final group presentations competences
- 14 16/12/2021 9:15-18:00 Group presentations To receive peer feedback on the proposition resulting from the Fieldlab and receive input for the final paper All team Presentations of 15-20 minutes per group and feedback session Group presentation (powerpoint or other tools), organisation of the session and division of tasks, prepping presenters 15% Group
- 15 Individual Paper Deadline 15/1/2022 23.59 45% Individual
5% Individual presence and participation
100%