

## INFOA EXCELIN SISÄLLÖSTÄ

INDI - Indikaattoreista ilmiöihin -hanke

Hankkeen työpaketti 2.1 Kartoitetaan mahdolliset digitaaliset vaihtoehdot demoa varten

Hankesuunnitelmasta:

Tunnistetaan digitaaliset alustat/ratkaisut joita voi käyttää seuraavien tavoitteiden saavuttamiseen

Digitaalisen ratkaisun tavoite (goal)

Laadullisen ilmiöohjaisen tiedon keruu, jakaminen, analysointi, päätöksenteko.

Mikä ilmiö? Ketkä ovat käyttäjiä? Minkätyyppisellä laadullisella datalla ilmiötä voi kuvata?

Digitaalinen ratkaisu/alusta toteuttaa yhden tai useamman toiminnon:

- |    |                     |   |
|----|---------------------|---|
| 1. | datan keruu         |   |
| 2. | datan säilytys      | * |
| 3. | datan valmistelu    | * |
| 4. | datan jakaminen     |   |
| 5. | datan analysointi   |   |
| 6. | datan visualisointi | * |
| 7. | tietoturva          | * |
| 8. | tietosuojat         | * |

\* ei varsinaisesti mainittu hankesuunnitelmassa

SHEET TOOLS 1 Työkaluja tekoälypohjaisten ratkaisujen rakentamiseen. Lista ohjelmistokehittäjille.

Taulukkoon on merkitty työkalun toiminta alla olevan listan mukaisesti

data management

datan /(tiedostojen) hallintatyökalu

data refinery and cleansing

datan käsittelyn automatisointi, jotta sitä voidaan hyödyntää eri järjestelmien välillä

data analysis

data-analyysin työkalu

data visualisation

datan visualisoinnin työkalu

data integration and transformation

model building

tekoälymallin kouluttamisen työkalu

model deployment

tekoälymallin käyttöönoton työkalu

model monitoring and asses

tekoälymallin toiminnan valvonta ja arviointi työkalu

code asset management

tietokoneen ohjelmiin liittyvien tiedostojen hallintatyökalu

development environment

ohjelmistokoodin kehitysympäristö

execution environment

ohjelmiston tuotanto/käyttöympäristö

SHEET TOOLS 2

Muita työkaluja, jotka voivat olla tukena datan käsittelyssä tai valmistelussa. Ohjelmistokehittäjille

Taulukkoon on merkitty työkalu ja sen käyttötarkoitus

TEXT-EXTRACTOR

IMAGE-EXTRACTOR

SHEET TOOLS 3

Työkaluja data-analyysiin. Tutkijoille.

QUAL DATA

QUANT DATA

SHEET TOOLS 4

Alustoja datan hyödyntämiseen.

TESTBED

DATASET

TOOLS 1		For software developers to build fully fledged digital solutions					
Interest	Name	URL	Type	Description	Scope	Last checked	Notes on onboarding / documentation
	MySQL	<a href="#">MySQL</a>	Open source	relational DB	data management	28.11.2023	
	PostgreSQL	<a href="#">PostgreSQL: The world's mos</a>	Open source	relational DB	data management	28.11.2023	
	MongoDB	<a href="#">MongoDB: The Developer D</a>	Open source	noSQL DB	data management	28.11.2023	
	Apache CouchDB	<a href="#">Apache CouchDB</a>	Open source	noSQL DB	data management	28.11.2023	
	Apache Cassandra	<a href="#">https://cassandra.apache.org</a>	Open source	noSQL DB	data management	28.11.2023	
	Apache Hadoop File System	<a href="#">Apache Hadoop</a>	Open source	file-based storage	data management	28.11.2023	
	Ceph Cloud File System	<a href="#">https://ceph.io/</a>	Open source	file-based storage; distributed storage	data management	28.11.2023	
x	Elasticsearch	<a href="#">Elasticsearch Platform — E</a>	Open source	text data and search index. Search and analytics engine.	data management	5.1.2024	Can be run on local computer. Instructions: <a href="https://www.elastic.co/guide/en/elasticsearch/reference/current/run-elasticsearch-locally.html">https://www.elastic.co/guide/en/elasticsearch/reference/current/run-elasticsearch-locally.html</a>
	Oracle DB		Commercial	industry-standard DB, commercial support	data management	7.12.2023	
	Microsoft SQL Server		Commercial	industry-standard DB, commercial support	data management	7.12.2023	
	IBM Db2		Commercial	industry-standard DB, commercial support	data management	7.12.2023	
	Apache AirFlow		Open source	Platform for programmatically authoring, scheduling and r	data refinery and cleansing	28.11.2023	
	Kubeflow		Open source	Machine learning toolkit on top of Kubernetes. It is a platfo	data refinery and cleansing	28.11.2023	
	Apache Kafka		Open source	Distributed streaming platform. Allows applications to publi	data refinery and cleansing	5.1.2024	
	Apache NiFi		Open source	Allows users to automate the flow of data between system	data refinery and cleansing	5.1.2024	
	Apache SparkSQL		Open source	SPARK ecosystem module for working with structured data	data refinery and cleansing	5.1.2024	
	NodeRED		Open source	visual editor that can combine hardware devices, APIs, and	data refinery and cleansing	5.1.2024	
	Apache Superset	<a href="#">Welcome   Superset (apach</a>	Open source	quantitative data handling. Web-based application to visu	data visualisation	5.1.2024	
	PlotlyDash		Open source	Library for creating interactive, exploratory data visualisat	data visualisation	5.1.2024	
	Tableau		Commercial	visualisation for end-user	data visualisation	5.1.2024	
	Microsoft PowerBI		Commercial	visualisation for end-user	data visualisation	7.12.2023	
	IBM Cognos Analytics		Commercial	visualisation for end-user	data visualisation	7.12.2023	
	Kibana		Open source	web-based visualisation tool. Used with Elasticsearch to an	data visualisation	5.1.2024	
	Hue		Open source	Apache Hadoop large dataset analysis and visualisation too	data visualisation	5.1.2024	
R			Open source	statistical data handling and visualisations.	data visualisation	5.1.2024	
	RODA	<a href="#">Welcome to RODA Project.</a>	Open source	qualitative data, integrates with R, supports plain text onl	data visualisation	5.1.2024	
	Apache PredictionIO		Open source	Machine learning server. Supports Apache Spark ML model	model deployment	5.1.2024	
	Seldon		Open source	supports nearly every framework including, TensorFlow, Ap	model deployment	30.11.2023	
	Milago		Open source	Library for serialising and deserialising learning models. Su	model deployment	5.1.2024	
	TensorFlow Lite		Open source	supports running TensorFlow machine learning models on r	model deployment	30.11.2023	
	TensorFlow.js		Open source	Library for building and deploying machine learning models	on the web in JavaScript. User can train and execute models	in the browser of on Node.js. Supports several model architectures: neural networks, decision trees, and k nearest neighbours.	
	TensorFlow Serving		Open source	System for machine learning models for production environm	ents. Supports TensorFlow models		
	Kubernetes / K8s	<a href="#">https://kubernetes.io/</a>	Open source	Platform to support launching, scaling and managing other	applications i.e. is a container framework running on a	5.1.2024	
x	Simple ML for Sheets	<a href="#">https://simplemlforsheets.com/</a>	Free	Simple ML for Sheets allows anyone to use machine learni	ng model deployment	5.1.2024	
	ModelDB		Open source	DB for machine learning models. Supports Apache Spark M	odel monitoring and assessment	30.11.2023	
	Prometheus		Open source	model performance measuring: accuracy, model bias	model monitoring and assessment	30.11.2023	
	IBM AI Fairness 360		Open source	detects and diminishes bias in ML models.	model monitoring and assessment	30.11.2023	
	IBM Adversarial Robustness 360 Toolbox		Open source	makes ML model more robust against adversarial attacks v	model monitoring and assessment	30.11.2023	
	IBM AI Explainability 360		Open source	makes machine learning models more transparent. Finds si	model monitoring and assessment	30.11.2023	
	git		Open source	code version control	code asset management	30.11.2023	
	GitHub		Open source	code version control	code asset management	30.11.2023	
	GitLab		Open source	code version control, can be hosted and managed on your	code asset management	30.11.2023	
	Atlassian Bitbucket	<a href="#">Atlassian Bitbucket Git Cod</a>	Freemium	code version control	code asset management	30.11.2023	
	Apache Atlas		Open source	data version and annotation tool	data asset management	30.11.2023	
	Linux Foundation odpi Egeria	<a href="#">https://egeria-project.org/</a>	Open source	platform that offers automatic capturing, managing and d	ata asset management	30.11.2023	
	Kylo	<a href="#">Kylo</a>	Open source	data lake management software platform with integrated r	data asset management	30.11.2023	
	Informatica Enterprise Data Management		Commercial	data governance, versioning and annotation	data asset management	7.12.2023	
	IBM InfoSphere Information Governance Catalog		Commercial	data governance, versioning and annotation	data asset management	7.12.2023	
	Jupyter	<a href="#">https://jupyter.org/</a>	Open source	Jupyter is a project that builds tools, standards, and servic	e development environment	5.1.2024	Kokellu <a href="https://docs.jupyter.org/en/latest/start/index.html">https://docs.jupyter.org/en/latest/start/index.html</a>
	JupyterNotebook	<a href="#">jupyter notebook document</a>	Open source	Original web application for sharing Jupyter computational	e development environment	5.1.2024	Project Jupyter   Installing Jupyter
x	JupyterLab	<a href="#">https://jupyterlab.readthedo</a>	Open source	highly extensible, feature-rich notebook authoring applica	t development environment	5.1.2024	Project Jupyter   Try Jupyter
	JupyterHub		Open source	Enterprise version of JupyterLab		5.1.2024	
	Apache Zeppelin	<a href="#">https://zeppelin.apache.org/</a>	Open source	similar to Jupyter Notebooks. Includes integrated plotting	e development environment.	7.12.2023	
	Anaconda / Anaconda Navia	<a href="#">https://www.anaconda.com</a>	Freemium	AI and data science platform. Anaconda Navigator	desktop application is used to manage integrated	applications, packages, and environments. Includes Python and R. Jupyter Notebooks can be used in the environment. Includes a large collection of libraries.	
x	RStudio		Open source	RStudio unifies programming, execution, debugging, rem	e development environment	5.1.2024	<a href="https://www.anaconda.com/download">https://www.anaconda.com/download</a>
	Spyder		Open source	Python-based IDE includes tools for advanced editing, anal	ysis, debugging, profiling, and visualisation.	5.1.2024	
	Apache Spark		Open source	cluster of servers can linearly scale the performance of	the	7.12.2023	
	Apache Flink		Open source	stream processing for real-time data streams	execution environment	7.12.2023	
	Risobal Ray		Open source	large scale deep learning model training	execution environment	7.12.2023	
x	Knime Analytics Platform	<a href="#">https://www.knime.com/</a>	Open source	fully integrated visual tool, can be extended by program	mi	7.12.2023	<a href="https://hub.knime.com/3oime5oaxes/Education?course=3bcw6640t4uag&amp;">https://hub.knime.com/3oime5oaxes/Education?course=3bcw6640t4uag&amp;</a>
	Orange		Open source	data integration and transformation, data visualisatio	n	7.12.2023	
	Informatica PowerCenter		Commercial	leading tool, visual interface	data integration and transformation	7.12.2023	
	IBM InfoSphere DataStage		Commercial	leading tool, visual interface	data integration and transformation	7.12.2023	
	SAP		Commercial	visual interface	data integration and transformation	7.12.2023	
	Oracle		Commercial	visual interface	data integration and transformation	7.12.2023	
	SAS		Commercial	visual interface	data integration and transformation	7.12.2023	
	Talend		Commercial	visual interface	data integration and transformation	7.12.2023	
	IBM Watson Studio Desktop		Commercial	visual interface, data visualisation for expert, includes	data	7.12.2023	
	IBM SPSS Modeller		Commercial	data mining tool, can export model in open source form	at	7.12.2023	
	SAS enterprise miner		Commercial	data mining tool	model building	7.12.2023	
x	Leximancer	<a href="#">https://www.leximancer.com</a>	Commercial	Text mining software for the analysis of the content of col	lect	11.12.2023	
	IBM Collaboration and deployment services		Commercial		model building, model deployment		

TOOLS 2

Other tools for software developers

Interesting	Name	URL	Type	Description	Scope	Last checked	Notes on onboarding / documentation
x	Xpdf	<a href="http://www.xpdf">http://www.xpdf</a>	General Public License (GPL), version 2 or 3	Xpdf is an open source viewer for Portable Document Format (PDF) files. The Xpdf project also includes a PDF text extractor, PDF-to-PostScript converter, and various other utilities.	TEXT-EXTRAC1	3.6.2024	pdftops, pdftotext, pdftohtml, pdffinfo, pdffonts, pdfdetach,

TOOLS 3 Tools for data-analysis

Interesting	Name	URL	Type	Descriptor	Scope	Last checked	Notes on onboarding / documentation
	NVivo	<a href="https://lumivero.com/products/">https://lumivero.com/products/</a>	Licence	NVivon käy	QUAL DATA		
	Atlas.ti	<a href="https://atlasti.com/">https://atlasti.com/</a>	Licence	Samankalt:	QUAL DATA		
	RQDA	<a href="https://rqda.r-forge.r-project.org/">https://rqda.r-forge.r-project.org/</a>	Open source	Laadulliseen	QUAL DATA		Tekstimuotoiseen dataan. SQL tietokanta vaaditaan.
	VosViewer	<a href="https://www.vosviewer.com/">https://www.vosviewer.com/</a>	Open source	Constructing and visualising	bibliometric networks		
	Elicit	<a href="https://elicit.com/">https://elicit.com/</a>	Freemium	Summarising papers			AI based tool
	Litmaps	<a href="https://www.litmaps.com/">https://www.litmaps.com/</a>	Freemium	Creates bibliometric map of	literature		AI based tool
	ResearchRabbit	<a href="https://www.researchrabbit.ai/">https://www.researchrabbit.ai/</a>	Free	discover and visualize relevant literature			AI based tool
	SemanticScholar	<a href="https://www.semanticscholar.org/">https://www.semanticscholar.org/</a>	Freemium	Search and organise research papers			AI based tool

#### TOOLS 4

Interesting Name	URL	Type	Description	Scope	Last checked	Notes on onboarding / documentation
Virtual Finland T	<a href="https://testbed.fi/">https://testbed.fi/</a>		Finnish dat	TESTBED	14.3.2024	
Open data	<a href="https://www.avoindata.fi/">https://www.avoindata.fi/</a>			DATASET		