WELCOME

SESSION 3

SPONSORED BY



FAIR DATA STEWARDSHIP AWARENESS COURSE

FAIR DATASTEWARDSHIP: A NEW PROFESSION





THE EMERGING DEMAND FOR HIGH QUALITY FAIR SERVICES

FAIR DATA STEWARD: A NEW PROFESSION

THE DATA PROBLEM IN RESEARCH AND INNOVATION

- Most data do not TALK to each other
- Data are lost and/or hard to find
- Constrains scaling of effective knowledge discovery
- Limits the delivery of a fully effective discovery and R&D
- 20% links to supplementary data 'rot away' (annually)
- 60 80% of data is lost forever
- 60% of what is literature can not be found by machines
- Only 12% of NIH funded datasets are deposited in recognized repositories
- Approximately 50% of funded research not reproducible

CAPACITY CHALLENGE FAIR DATA STEWARDS





Ever growing Big Data Tsunami requires skill sets beyond today's computer science education

• Need for education programs for professional Data Stewards

- EC: 1.7M scientists and 70-100M professionals in S&T need 500,000 linked data stewards (based on conservatively assuming 10M data-producers and 1 data-steward per 20 data 'generators').
- USA: "The United States of America faces a shortage of 140,000 to 190,000 with analytical expertise and 1.5M managers and analysts with the skills to understand and make decisions based on the analysis of Big Data" (Source McKinsey Global Institute, 2011)

JUST AN IMPRESSION OF HOW BIG THIS MARKET IS

- €2B for initial phase EOSC
- Total EU plus USA 86,5 B for Data Stewardship (DS) annually
- EU (28 member states)
 - GDP 52,000 B
 - 2.4 % of GDP to R&D = 1,248 B
- The Netherlands
 - **GDP 818 B**
 - 1,973% of GDP to R&D = 16B

(Source OECD)





EXAMPLES OF ORGANIZATIONS IN THE PROCESS OF GOING FAIR



















































- Many organizations (20+) have participated in FAIR BYODs and trainings
- Several academic institutions and funders have started or are considering GO FAIR Readiness programs
- Several companies have started or are considering the GO FAIR Readiness program

FAIR DATA STEWARDSHIP: NO LONGER WHY BUT HOW

THE CAPACITY CHALLENGE

CAPACITY CHALLENGE FAIR DATA STEWARDS

• Only a few small size companies/institutions (globally) with sufficient knowledge to:

Train FAIR Data Stewards



• Make data FAIR



Create FAIR tooling



Assist with FAIR Data Stewardship Plans

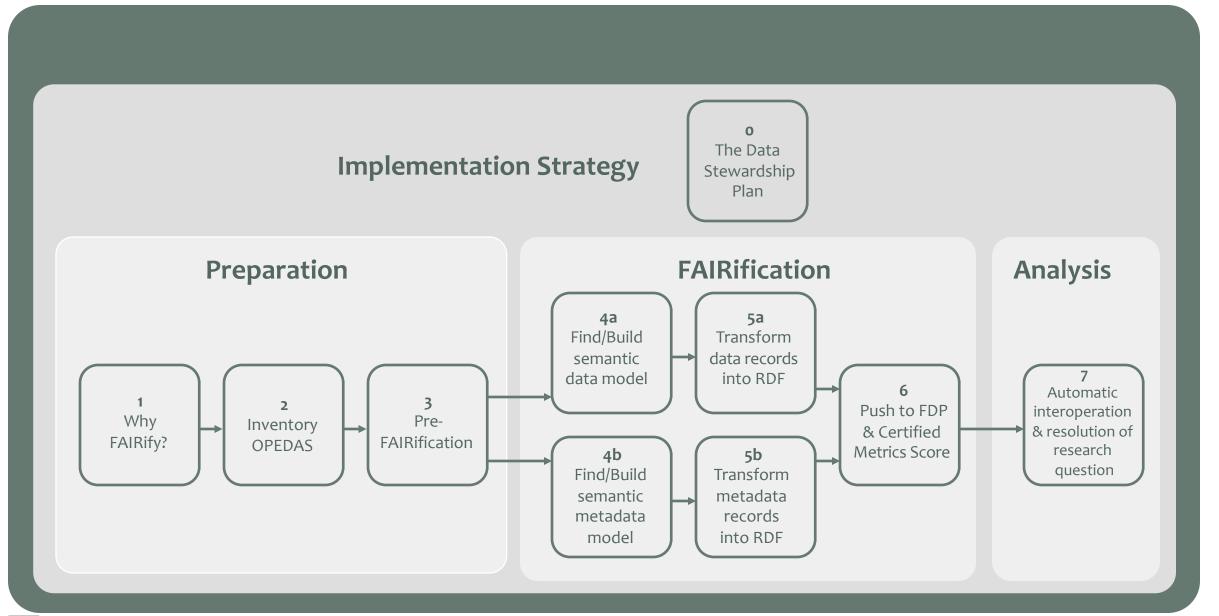


Assist organizations/companies to GO FAIR



• This means an *enormous* economic activity and *lots* of new jobs

THE 7 CANONICAL STEPS OF FAIRIFICATION



FAIR DATA SUITE OF TOOLS











FAIR Metrics Evaluator Quantitative | Reproducible | Objective

PROBLEM (WITH ALL DUE RESPECT) **ALL TOOLS ARE CURRENTLY PROFESSORWARE**







THE FAIR SERVICE PROVIDER CONSORTIUM

FROM PROFESSORWARE TO PROFESSIONAL WARE

THE GO FAIR FOUNDATION

- Founded February 2018
- Goal: Legal entity to support the GO FAIR Initiative towards implementation
- Image: Advisory, non-competitive, non-profit
- Focus: Broker function for early market initiatives: Grant proposals, Pilots, Training, Hackathons, etc.

GOOD Certification:

- Professional FAIR Data Stewards
- Datasets
- Goodling
- Service Providers
- Organisations/companies



THE INITIATIVE



The GO FAIR Foundation in collaboration with Phortos Consultants, a DTL partner, has taken the initiative to approach candidates for the GO FAIR Service Provider Consortium



GO FAIR SERVICE PROVIDER PARTNERS CONSORTIUM: covering the full spectrum of FAIR Data and Services







Data Quality



FAIRification and Analysis





Consulting and training



Digital Engagement



Certification and Coordination





Software development and consulting



FAIR data at the source







GO FAIR SERVICE PROVIDER PARTNERS CONSORTIUM ADHERE TO

- •• The GO FAIR Rules of Engagement
- The GO FAIR Vision
- •• The GO FAIR Readiness Program approach (interim)
- •• The GO FAIR Certification rules (soon to come)
 - The GO FAIR Implementation choices
 - The GO FAIR FAIR Data Stewardship approach
 - The GO FAIR training and tooling approach



SPEAKING WITH ONE VOICE

THE GO FAIR SERVICES PROVIDER CONSORTIUM'S OFFERINGS

• FAIR Data Consulting and Services

- FAIR Readiness Requirements studies
- FAIR Data Stewardship Planning
- FAIRification services
- Semantic and Ontology Modeling
- Analytics based upon integration of FAIR public and proprietary data
- FAIR Readiness training courses (off-site or on-site)
- **GOOD** FAIR Metrics Evaluation

GODE FAIR Value Events

- Demonstrates Value by answering driving research question making one of your datasets FAIR
- 6 weekly conference calls (one hour) in prep
- 2-day on site event: Bring Your Own Data (BYOD) and FAIR Metrics evaluation

FAIR Readiness Implementation and Vision Plan of Approach

- An introduction to GOing FAIR for organizations
- A 3 to 6-month implementation process including training and use cases





THE GO FAIR SERVICES PROVIDER CONSORTIUM'S TRAININGS OFFERING

FAIR TRAINING COURSES (on-site or off-site)

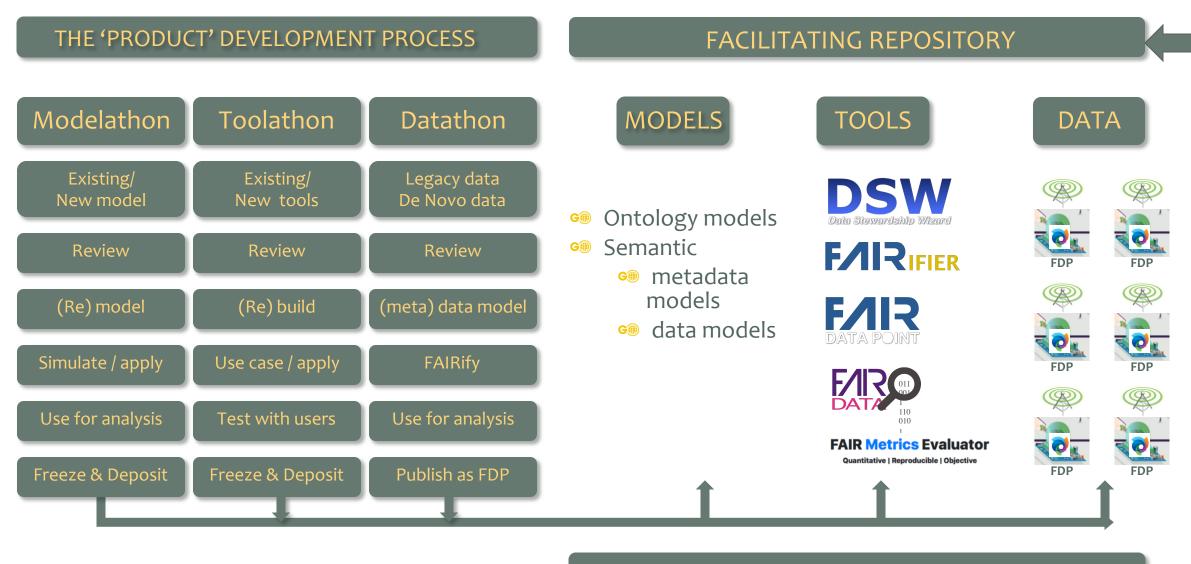
- G FAIR Awareness (one day)
- FAIR Data Stewardship (4 days; prerequisite Level 1)
- FAIR Ontology and Data modeling (4 days; prerequisite Level 1)
- FAIR Data & Services (4 days; prerequisite Level 1)
 - Operators: FAIR data processing
 - Engineers: FAIR tooling and apps

THE GO FAIR READINESS IMPLEMENTATION PROGRAM

- A 6 month project to make an organization GO FAIR Ready
 - Blueprint plan of approach including business case and design of the organization
 - Project initiation: clear roadmap for implementation
 - Project execution: Set up governance; set up structures for frontrunners working on (3) use cases
 - Involve relevant stakeholder from the beginning to show value
 - National and international engagement and collaboration

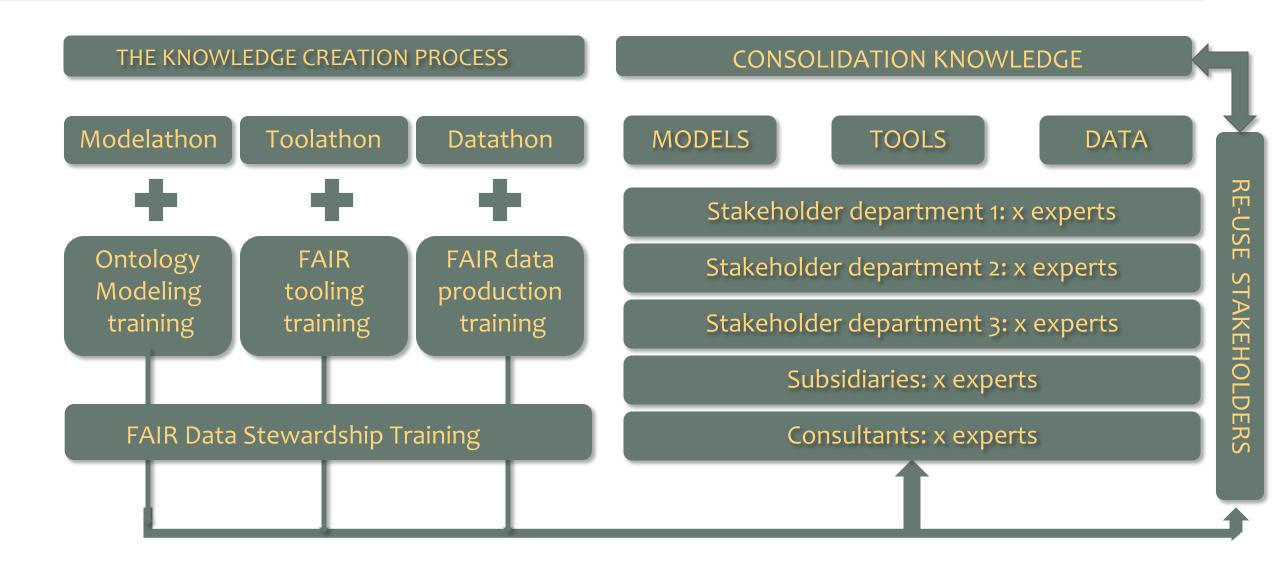
RE-USE OTHER STAKEHOLRDERS

Stakeholder driven implementation and safeguarding knowledge



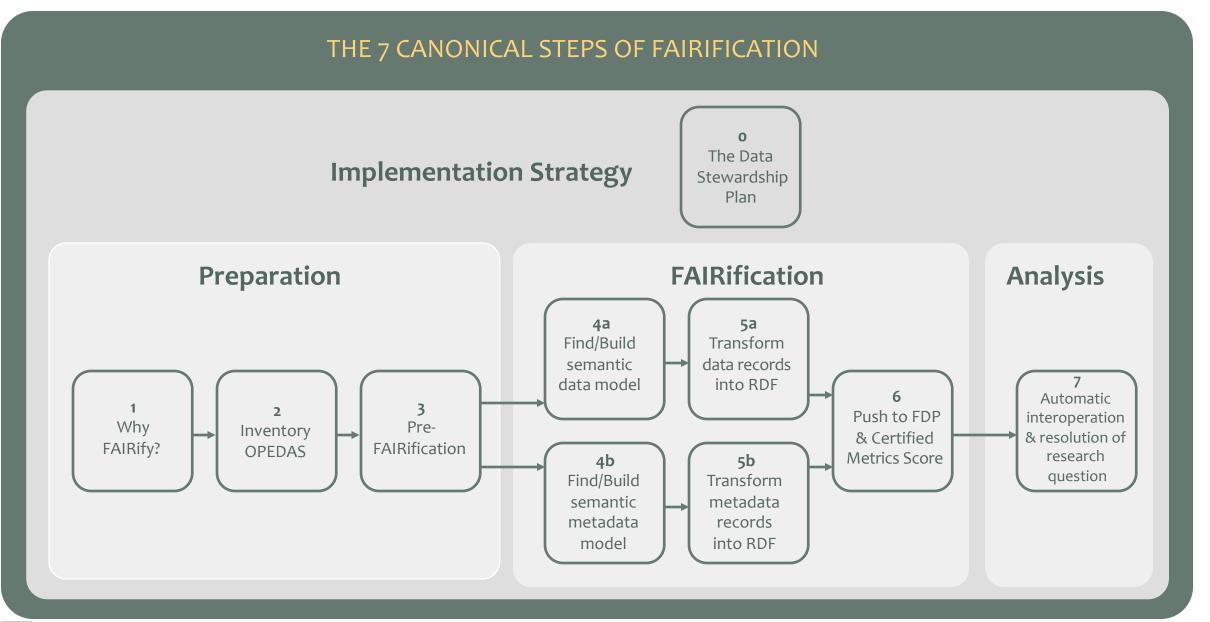
AVAILABLE AS ONLINE PROFESSIONAL SERVICES

SAFEGUARDING KNOWLEDGE









DAY 1

GO FAIR: an introduction

- ✓ Introduction and purpose of course
- ✓ The need for FAIR data
- ✓ The history of the FAIR initiative
- ✓ The internet of FAIR Data and Services

An introduction to FAIR Data Stewardship

- ✓ What is FAIR Data Stewardship
- ✓ The purpose and goals
- ✓ The FAIR Principles and Metrics

Trainers

Erik Schultes Albert Mons

Lunch

FAIR Data Stewardship: a new profession

- ✓ The need for high quality FAIR services
- ✓ Elements of a FAIR Data Stewardship Department
- ✓ Roles in a FAIR Data Stewardship Department
- ✓ A FAIR Readiness Implementation Program

Practicing FAIR Data

- ✓ The FAIR Principles explained
- ✓ The FAIR Metrics applied
- ✓ The FAIR Community Challenges discussed
- ✓ Resources for FAIR Data Stewards

Erik Schultes Albert Mons

Introduction to Semantic Data Modeling and Ontologies

- ✓ What is semantic interoperability
- √ How can it improve the current data situation
- ✓ Ontological principles
- ✓ Ontologies are computer-actionable artefacts

Trainer

Luiz Bonino

Lunch

Introduction to Semantic Web and Linked Data

- ✓ The Semantic Web
- ✓ Linked Data
- ✓ Unique Identifiers
- √ The FAIR principles explained
- \checkmark For each Principle what are the required actions

Luiz Bonino

DAY 3

FAIRification Process - DATA

- ✓ Where will the data "live" (repository)? Considerations for GUPIDs for the FAIR data
- ✓ Consider semantic models for the sample dataset
- ✓ Consider "core" ontological frameworks (e.g. SIO, DCAT)
- ✓ Apply semantic model to data elements
- ✓ Custom scripts to achieve data record transformation to FAIR Data
- ✓ "Push" into the selected repository for data and metadata) = machine-readable knowledge graph

Trainers

Mark Wilkinson/ Luiz Bonino

Lunch

FAIRification process

- ✓ Open Refine FAIRifier tool
- ✓ Sculpting and cleaning data
- ✓ Export data without custom scripting

Mark Wilkinson/ Luiz Bonino

DAY 4

FAIRification process - METADATA

- ✓ The purpose of metadata
- ✓ What should be included to be FAIR
- ✓ Meta data structures (FAIR Accessors and LDP containers)
- ✓ Introduction to the FDP
- ✓ Create a FAIR metadata record by scripting
- ✓ Push into a FAIR metadata repository

Trainers

Mark Wilkinson/ Luiz Bonino

Lunch

The FDP in practice

- ✓ Create FDP record for the data set published on Day 3
- ✓ Explore the FDP (interface and SPARQL)
- ✓ Are we FAIR yet: discuss measuring FAIRness
- ✓ The FAIR metrics
- ✓ Metric Evaluator (prototype)
- ✓ Evaluate published data / metadata

Mark Wilkinson/ Luiz Bonino

DAY 5

Answering driving questions: the real power of FAIR data

- ✓ Query federation
- ✓ More SPARQL if needed
- ✓ Through a search or analysis demonstrate value
- ✓ Discussion: what did we achieve?
- ✓ Professional Analytics Tool: an example
 - ✓ The Euretos AI Platform

Trainers

Mark Wilkinson/ Albert Mons

Lunch

Practical Application and next steps

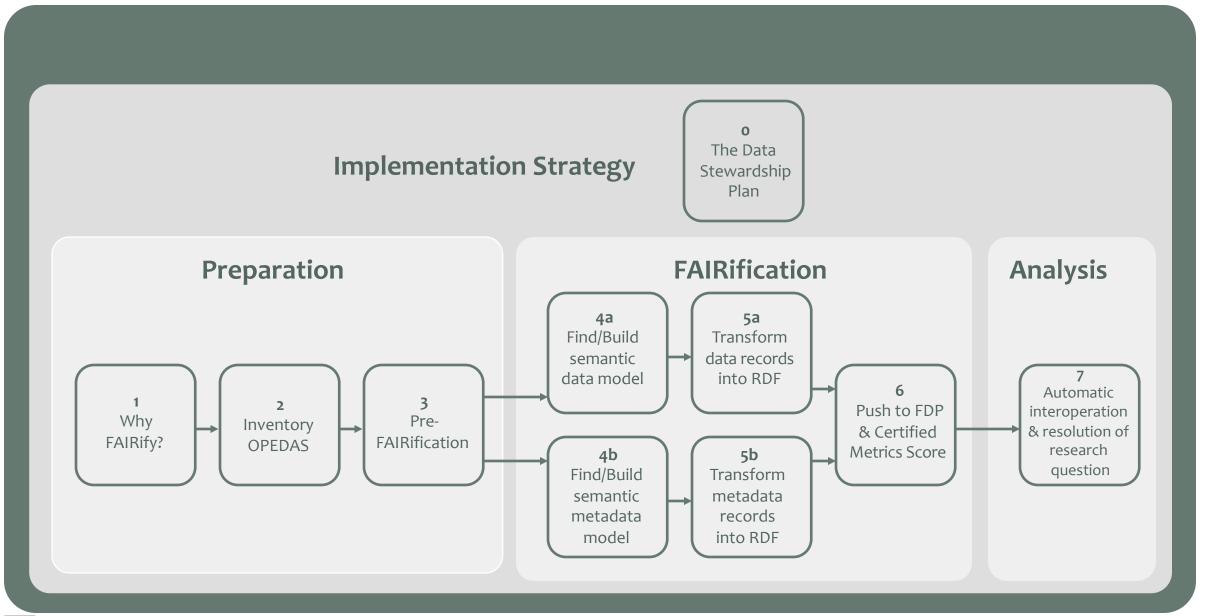
- ✓ Scenarios and Plan of Action for follow up
- ✓ Institutional FAIR Data Stewardship
- ✓ Plans for Train-the-trainers

Erik Schultes Albert Mons





THE 7 CANONICAL STEPS OF FAIRIFICATION



Introduction to FAIR Data and GO FAIR

- ✓ Introduction and purpose of course
- ✓ The need for FAIR data
- ✓ The history of the FAIR initiative
- ✓ The internet of FAIR Data and Services
- ✓ Data stewardship

Trainers

Albert Mons

Lunch

Introduction to ontology-driven conceptual modeling – I

✓ The relation between Conceptualization, Language, (Meta)Models and Ontology

Introduction to ontology-driven conceptual modeling – II

✓ The criteria for an Ontologically Well-Founded Conceptual Modeling Language

Giancarlo Guizzardi

Types and taxonomic structures

- ✓ Object Type Categories and Taxonomic Structures
- ✓ A typology of Categories of Object Types: Philosophical and Psychological

Trainer

Giancarlo Guiizzardi

Lunch

Types and taxonomic structures

Continued

- ✓ Object Type Categories and Taxonomic Structures
- ✓ Ontology Patterns for Modeling of Taxonomic Structures

Giancarlo Guiizzardi

Relationships and events

- ✓ Ontological Analysis and Modeling of Relations
 - ✓ A Typology of Relations
- ✓ Relationship Reification and Truthmaking Patterns
 - ✓ An Ontology of Events with Applications

Trainer

Giancarlo Guizzardi

Lunch

From the Conceptual to the Operational Level

- ✓ Modeling session
- √ Validation and Anti-Pattern Detection
- ✓ Aspects of mapping to implementation environments
- ✓ Wrap up of ontology-driven conceptual modeling

Giancarlo Guizzardi

Conceptual models through ontologies on the semantic webs

- ✓ Introduction to the Semantic Web
- ✓ Introduction to Linked Data

Trainer

Luiz Bonino

Lunch

Conceptual models through ontologies on the semantic web

- ✓ Practical data modeling and data manipulation using the Semantic Web
- ✓ Relating linked data using ontologies

Luiz Bonino

Linkable Metadata modeling

- √ Hands-on:
 - ✓ Defining required metadata elements
 - ✓ Select existing vocabularies for the concepts of the metadata content
 - ✓ Create metadata records based on the defined templates

Trainer

Luiz Bonino

Lunch

FAIR Data Points

- ✓ FAIR Data Point and metadata layers
- ✓ Deploying FAIR Data Point
- ✓ Populate FDP with metadata
- ✓ Link FDP metadata to the data

Luiz Bonino



IMPLEMENTING FAIR DATA STEWARDSHIP

- Board level decision to GO FAIR
- Formulate a institution wide Data stewardship policy
- Set goals for how to become more FAIR
- Formulate an implementation program
- Determine and approve adequate budgets

FAIR DATA STEWARDSHIP FOR ORGANIZATIONS

- Implementing FAIR Data Stewardship organization-wide
- Elements of a Department of Data Stewardship
- Roles in a Department of Data Stewardship
- The FAIR Service Provider Consortium Partners
- **SET OF SET OF S**

ELEMENTS OF A DEPARTMENT OF DATA STEWARDSHIP (1)

- Well embedded (thus findable and trusted) in its organization
- Well run and organized at a supra-department level (hub and spokes model).
- Good Formulate an institution wide Data stewardship polic,
- Overseeing all vital Data Stewardship resources
- The place to go to find Data Stewardship Experts who can support

GO FAIR Data Stewardship Team



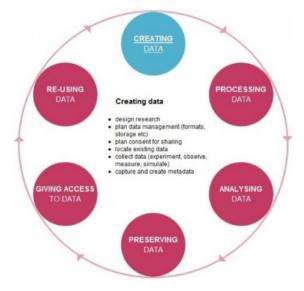
ELEMENTS OF A DEPARTMENT OF DATA STEWARDSHIP (2)

Regard Data Stewards as:

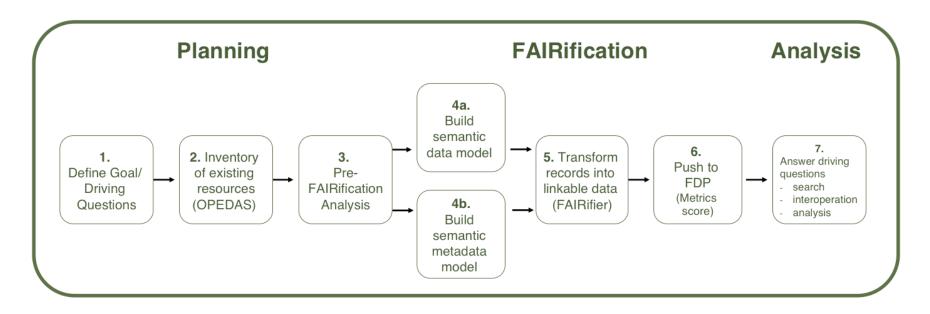
- Highly respected colleagueswith dedicated career tracks
- Real partners of the scientists not 'just data crunchers'.
- Involved
 - in the design of the research project
 - throughout the full research cycle.

GO FAIR Data Stewardship Team





ROLES IN A DEPARTMENT OF DATA STEWARDSHIP



- The FAIR Program manager
 - Oversees the end-to-end FAIR Readiness program
- The FAIR Data Steward
 - Oversees data `life cycles and use cases/projects
- The FAIR Data & Services Operator/ Engineer
 - Operates tooling and actively makes data FAIR
 - Develops tooling & apps





ZORGINSTITUUT NEDERLAND PILOT

(HEALTH CARE INSTITUTE OF THE NETHERLANDS)

ZORG INSTITUUT NEDERLAND PILOT (HEALTH CARE INSTITUTE OF THE NETHERLANDS

Implement three use cases with the Personal Health Train.

• Intra Arterial Trombectomy

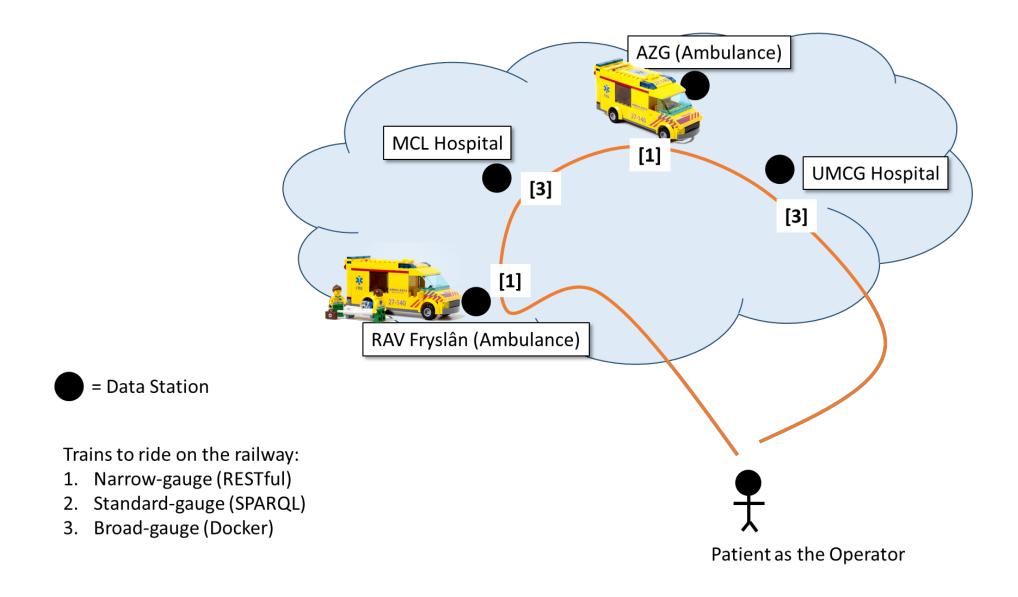
• Wet langdurige zorg (law on long term health care)

GO FAIRification on non-structured data

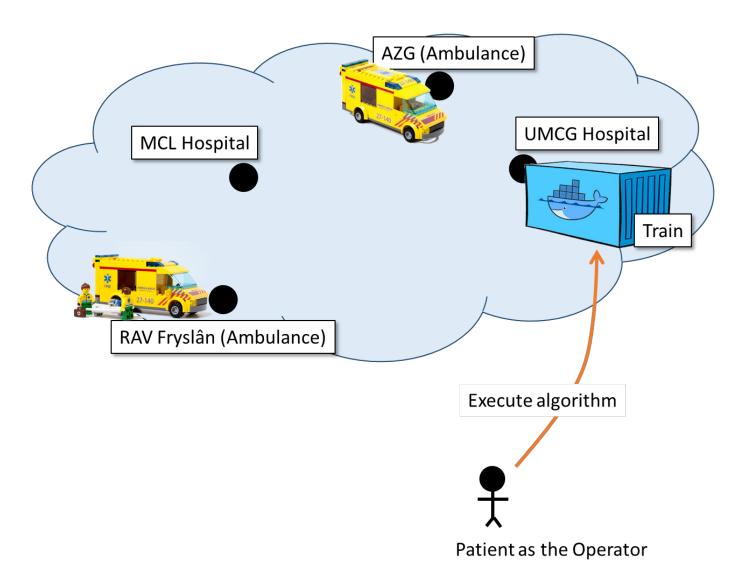
INTRA ARTERIAL TROMBECTOMY PILOT

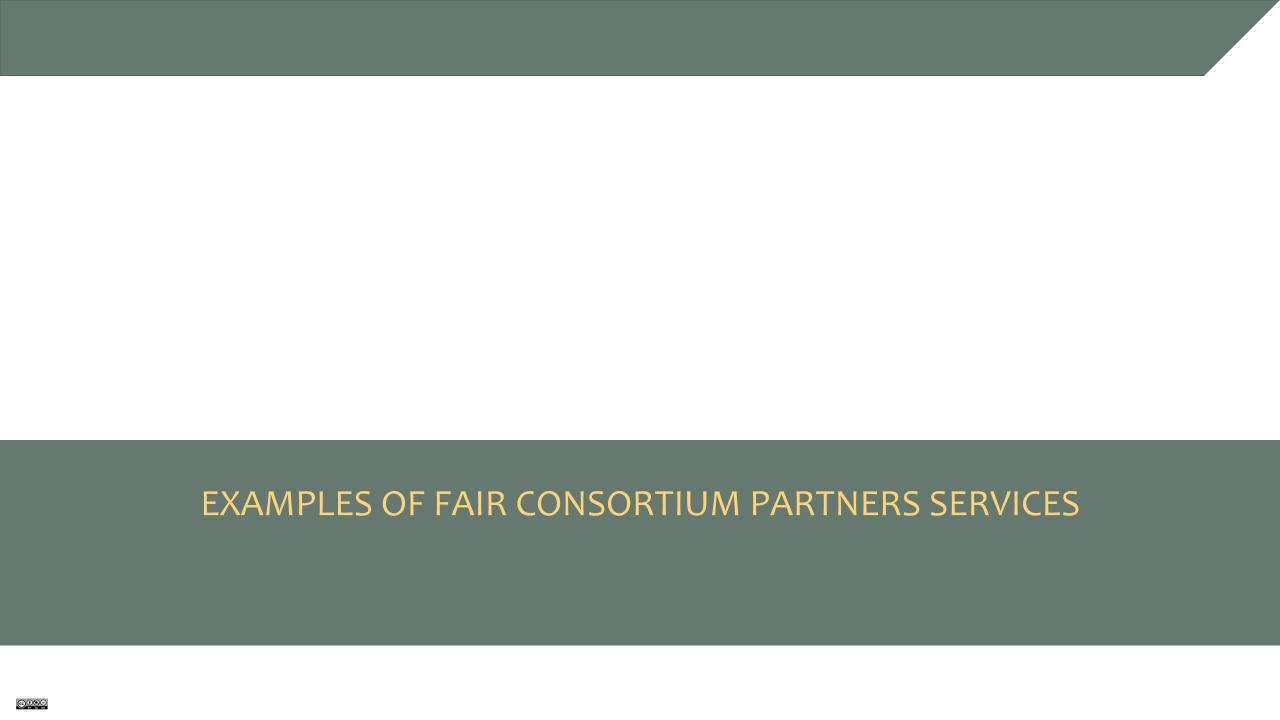


INTRA ARTERIAL TROMBECTOMY PILOT



INTRA ARTERIAL TROMBECTOMY PILOT







EXAMPLES OF FAIR CONSORTIUM PARTNERS SERVICES

FAIR GUIDELINES ASSOCIATED SERVICES



Which services do we offer around FAIR guidelines?

We are fully aware that most data is not either FAIR or not, but can exhibit several degrees of FAIRness. At The Hyve, we try to support you in taking the right measures to reach the highest level of FAIRness.

Services primarily aimed at data owner of exisiting data

- Assessment of the FAIRness of your data.
- Identification of ways to quickly and efficiently improve the FAIRness of your data!

Services primarily aimed at researchers that plan projects

- Creation of a data management plan compliant to FAIR guidelines.
- Assistance in grant proposal preparations.





EXAMPLES OF FAIR CONSORTIUM PARTNERS SERVICES

PROTOTYPE IN TESTING

FAIR Metrics Evaluator

Quantitative | Reproducible | Objective



Run the manual or semi-automated FAIR Metrics & discover how FAIR your data currently is.



Improve the FAIRness of your dataset and run the Evaluator again.
View the progress made.



Up to Date

New tools and Services will be available here. We are committed to helping you to become FAIR compliant.



Our mission

Our mission is to help making your data FAIR: Findable, Accessible, Interoperable & Reusable.

Your first step is to evaluate how FAIR your data set is

Find out how FAIR your dataset is here!

RUN EVALUATOR!

My metrics / Manual check / New metric

Manual check

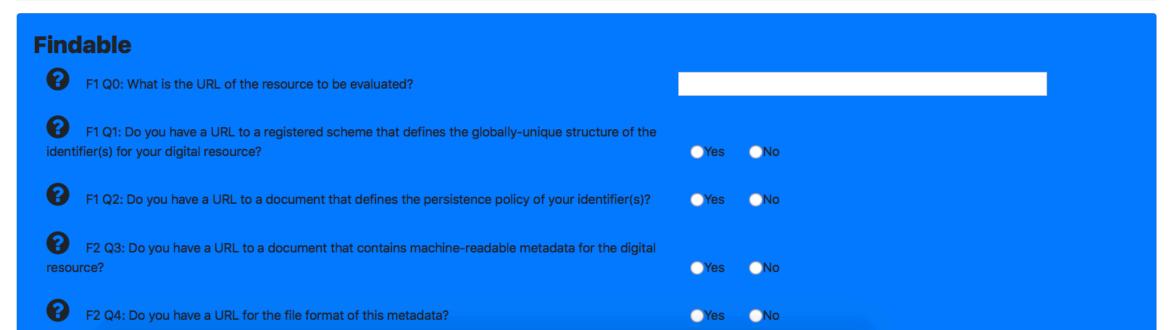
Would you like more information first? Please click here.

Select an earlier saved dataset evaluation or evaluate a new dataset

Create new dataset

Specify the name of your dataset evaluation:

Name your dataset here



Dataset: Test set 1

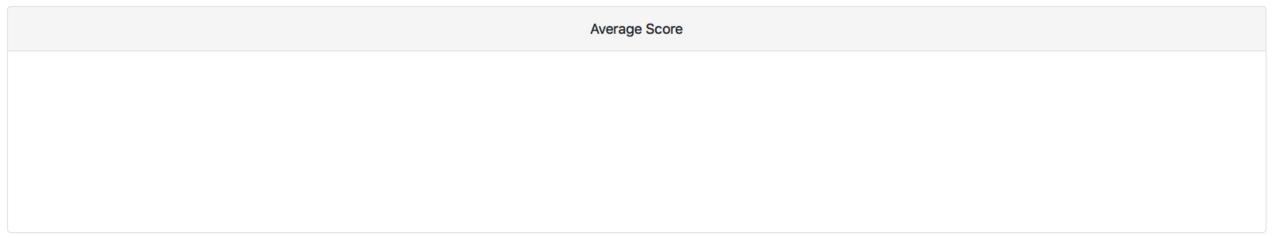
air metrics per run & the progression							
Date run	Evaluator	Findable	Accessible	Interoperable	Reusable	Score	
2018-05-16	Manual check	62%	16%	66%	66%	53%	

Questions answers of your last evaluation

Q0: What is the URL of the resource to be evaluated?	
Q1: Do you have a URL to a registered scheme that defines the globally-unique structure of the identifier(s) for your digital resource?	×
Q2: Do you have a URL to a document that defines the persistence policy of your identifier(s)?	٧
Q3: Do you have a URL to a document that contains machine-readable metadata for the digital resource?	٧
Q4: Do you have a URL for the file format of this metadata?	×
Q5: Do you have a URL to the metadata document that contains the globally unique and persistent identifier for the digital resource?	٧
Q6: Do you have a URL to the data described by in that metadata document?	٧
Q7: Is there a URL to a search engine that can find your data?	X

Metrics / My metrics

Your FAIR score



Findable

62%

Accessible

16%

Interoperable

66%

Reusable

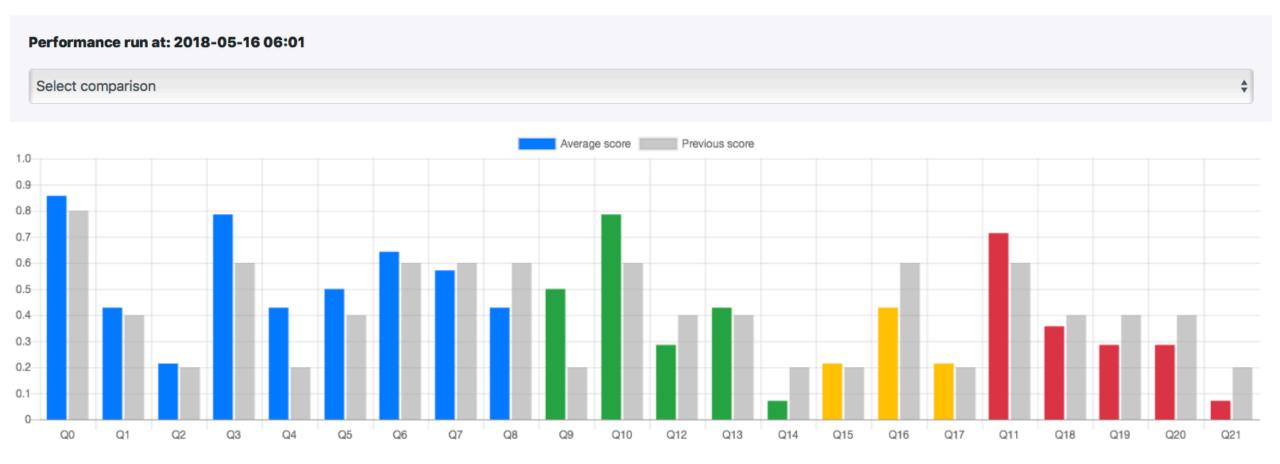
66%

Go to overview



Admin / Metric performances / Manual check / 2018-05-16 06:01

Metric evaluator: Manual check



EURETOS

EXAMPLES OF FAIR CONSORTIUM PARTNERS SERVICES

ELSEVIER CASE EXAMPLE – AI GENERATED HYPOTHESES

Joint pilot to enrich scientific publications with hypotheses that are generated using the Euretos AI Platform.

Focusing on an open question in an article's discussion section, a hypothesis is created and discussed in a Mendeley group.

The aim of the pilot is to assess whether these AI derived hypotheses can provide a catalyst for further scholarly discussion





How big data and AI can help you generate your scientific hypothesis

An Elsevier journal team works with Euretos to explore how machine learning and data analytics can guide research

By Valentina Sasselli, PhD and Hylke Koers, PhD February 2, 2018





EURETOS

Link to Euretos AI generated hypothesis:

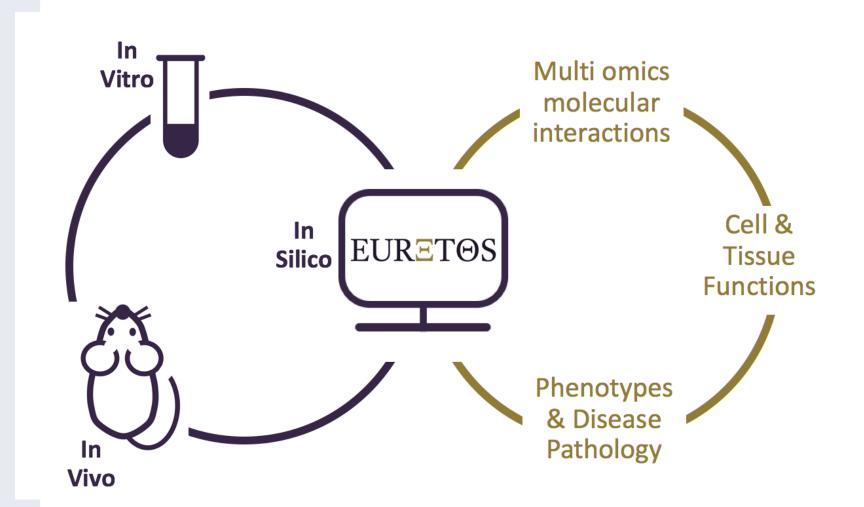
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3098734

AI DRIVEN SYSTEMS BIOLOGY APPROACH

Euretos puts the power of Al technology and big data analytics in the hands of the researcher via an easy-to-use front end.

By integrating all relevant multi-omics data as systems biology approach to biomarker and target research is enabled.

Researchers discover and evaluate how molecular mechanisms influence cell and tissue functions, and in turn mediate phenotypes and disease pathology.

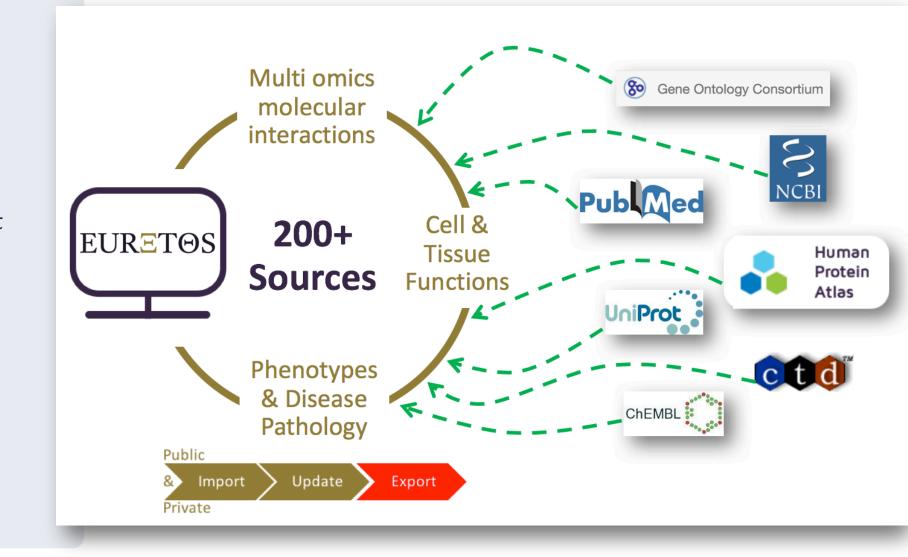




EURETOS AI PLATFORM – EXAMPLE OF INTEGRATED DATA APPROACH

By integrating over 200 public data sources, the platform provides the largest single environment with literature, experimental and clinical evidence

The multi omics data includes: genetic, genomic and proteomic annotations, expression profiles, experimental and animal models, diseases, phenotypes, pathways, small molecules - covering metabolites, food ingredients, as well as therapeutic agents (antibodies and peptides).





TYPES OF INTEGRATED DATA

Three types of data are loaded:

- 1. Ontologies & vocabularies
- 2. Structured data (multi-omics)
- 3. Unstructured data

All data can also be exported into a FAIR format where each individual data elements is ontologically resolvable enabling the data to be:

- Findable
- Accessible
- Interoperable
- Re-usable

- **1. Ontologies & vocabularies –** GO, PO, HPhO, MPhO, VTO, CMO, EFO, UMLS, MeSH, SNOMED etc.
- 2. Multi-omics Databases (structured data) Genomics & Genetic variation, Model organisms, Transcription regulation & expression, PTMs, Protein abundance, Protein Interactions, Enzymatic reactions, Metabolic interactions, Pathways, Experimental data
- Unstructured data Literature, Clinical Trials, Patents & RWD













ClinicalTrials.gov



TYPES OF ANALYTICS PROJECTS BASED ON INTEGRATED DATA

World leading pharma, biotech and academic institutions use the Euretos Al Platform to accelerate multi omics research in all major disease areas.

In addition to providing access to the AI platform, Euretos has undertaken over 50 projects in biomarker discovery, target identification, indication expansion, target validation and drug response & resistance.

EURΞTΘS

- Biomarker Discovery for all marker types: mechanistic, outcome, pharmacogenomic, toxicity biomarkers and diagnostic
- Target Identification for complex diseases, including upstream analysis of dysregulated transcripts and downstream impact
- Indication Expansion for known therapeutic agents for same tissue based pathologies as well as in non-related tissues and cell types
- Target Validation for specific phenotypes and pathologies including the comparison to similar targets, druggability and safety concerns
- Drug Response & Resistance analysis of molecular mechanism including the identification of adjuvant therapeutic interventions

Over **50 projects** in **all disease areas** including:

Pancreatic cancer, lung cancer, colorectal cancer, hematological tumors, complex cardiometabolic diseases, endometriosis, skin disorders, NASH, rheumatoid arthritis, inflammatory respiratory disease, pulmonary fibrosis, neuropathic pain.

WHAT'S IN IT FOR ME?

THE VALUE PROPOSITION

BENEFITS OF FAIR DATA STEWARDSHIP IF YOU ARE A FUNDER/POLICY MAKER

- Better organized stakeholder community
- Less resources lost on slack and overhead
- Increased 'Return on Investment' of public funding
- Automated FAIR Data Stewardship planning and FAIRness Evaluation
- Participate in international developments, e.g. European Open Science Cloud
- Improved societal impact:
 - on increased involvement of citizen/patient (digital control of own data
 - Increased economic benefits in health care sector including prevention

BENEFITS OF FAIR DATA STEWARDSHIP IF YOU ARE AN INSTITUTION

- No more silo-ed data 'solutions'
- Less time lost on data crunching; more time for research
- No more short term point solutions over and over again
- Compliance by design to technical, ELSI and scientific standards
- Easier & safer (inter)national exchange
- More efficient business operations

BENEFITS OF FAIR DATA STEWARDSHIP IF YOU ARE A CITIZEN

- Better opportunities for active participation
- Active recommendations for prevention
- With better prevention lower insurance premiums
- Better privacy: you are in control of your own data!
- More benefit from tax and charity money spent
- Faster development of new preventive, diagnostic and therapeutic solutions

BENEFITS OF FAIR DATA STEWARDSHIP IF YOU ARE A COMPANY OR ENTREPENEUR

- Have your researchers do research rather than data wrangling
- Offer your researchers better analytics, applications and services
- ALL your (proprietary) data
 - Integrated so you 'know what you know'
 - Compliant and interoperable with FAIR public domain data and EOSC procedures
 - Easier compliance with Medical Device Regulation and FDA Regulatory Procedures
 - Easier Post Implementation surveillance
 - Improved data analytics and knowledge predictions
 - Enabling more effective discovery process
 - Leading to decreasing Time-to-Market
 - Effecting your bottom line
- PR value: FAIR compliance and Open Science association
- **Save** cost and **increase** revenue

BENEFITS OF FAIR DATA STEWARDSHIP IF YOU ARE A PI/RESEARCHER

- Eligible for funding for FAIR Data Stewardship (the '5%')
- Assistance from dedicated professional FAIR Data Stewards
- More time for actual research!
- Your data interoperate with other data
- Increased scientific impact (citations; also of datasets!)
- Increased rate of hypotheses testing: Discovery and Innovation