Integrating AI into chemical safety assessment
Opportunities, challenges, and the path forward

October 16 2024
https://www.ecetoc.org/event/ai-workshop/

FAIR Data and data that are Fully AI Ready

Erik Schultes
http://orcid.org/0000-0001-8888-635X
eriks@gofair.foundation
These slides: https://osf.io/ueqvw

Integrating AI into chemical safety assessment
Opportunities, challenges, and the path forward

Opportunities

AI-assisted Next Generation Risk Assessment enabled by FAIR Data and Knowledge Graphs

Barry Hardy[1], Asmaa Ali[1], Tomaz Mohoric[1], Daniel Burgwinkel[1], Jeff Wiseman[1], Thomas Luechtefeld[2], Zaki Mughal[2]
[1] EdelweissConnect GmbH, Switzerland, [2] InSilica, Bethesda, MD 20817, US

SUMMARY

Our goal is to develop AI-assisted Next Generation Risk Assessment. In our first project phase we are developing processes and applications to prepare FAIR[1] datasets and knowledge graphs connecting compound information, biological response data, pathways and key events.

BACKGROUND

We developed two toxicology knowledge resources for case studies:
— A dataset and knowledge graph from the EU-ToxRisk[2] program on New Approach Methods in Toxicology.
— A knowledge graph based on a network model for steatosis, supporting ASPIS[3] case studies and SSbD4CheM.

TESTING AND REFINING

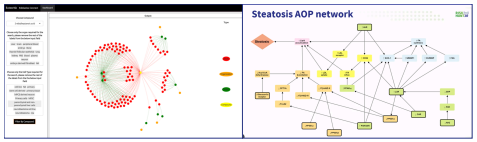
Our AI-enabled risk assessment test cases include:
1. Efficient retrieval of reliable, well-referenced knowledge via a knowledge graph;
2. Enhanced toxicology learning through ML and Generative AI models;
3. AI integration into risk assessment workflows for analysis, evidence interpretation, and decision-making;
4. Generation of auditable reports documenting risk assessment processes with supporting data and knowledge records.

NEXT STEPS AND COLLABORATION

Collaboration opportunities:
— sharing open science resources supporting NAMS-based risk assessment with the regulatory acceptance purpose.
— Establishing FAIR practices and approaches to sustainability.

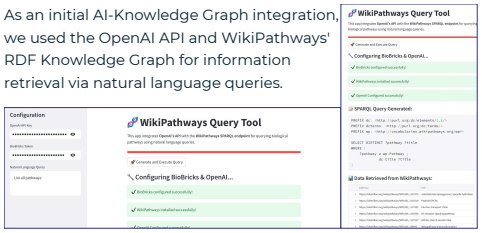
KNOWLEDGE GRAPH CREATION

The open knowledge resources offer a harmonized data template aligned with FAIR principles, omics data management best practices, and a standardized data collection approach focusing on harmonization, referencing, and integrity.



AI-ASSISTED APPLICATION

As an initial AI-Knowledge Graph integration, we used the OpenAI API and WikiPathways' RDF Knowledge Graph for information retrieval via natural language queries.

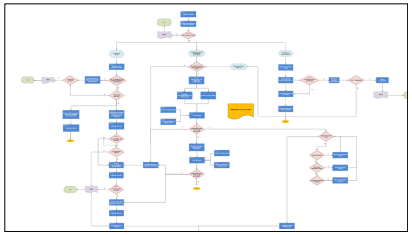


DESIGN FOR AI SAFETY (CBRN)

We are designing our toxicological safety risk assessment system to comply with AI safety regulations, such as the NIST AI Risk Management Framework and the EU AI Act. The providers of Large Language Models (LLMs) have to prevent the misuse in context of CBRN threats (Chemical, Biological, Radiological, and Nuclear). We analyse how these AI safety measures impact our use cases for safety risk assessments.

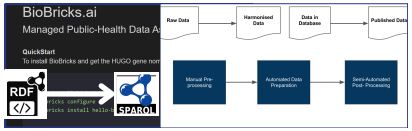
NGRA WORKFLOW INTEGRATION

Our AI-assisted knowledge from FAIR data is integral to NGRA workflows, including ASPA Workflow[3], supporting stages like problem formulation, hazard characterization, mechanistic interpretation, and points of departure.



QUALITY AND TRUST ISSUES

We design systematic, automated workflows for data processing and utilize the biobricks.ai platform to transform our FAIR datasets into FAIR knowledge graphs using the semantic web's RDF format.



CONTACT US
customersupport@edelweissconnect.com
edelweissconnect.com
Hochbergerstrasse 60c, 4057 Basel CH
Book a 1-to-1 demo scanning the QR code or learn more about our services from our website.
keep in touch soon this code

Led by: Edelweiss Connect

References
[1] Findable, Accessible, Interoperable, Reusable (FAIR)
[2] EU-ToxRisk, An Integrated European 'Flagship' Programme Driving Mechanism-based Toxicity Testing and Risk Assessment for the 21st century
[3] Animal-free Safety assessment of chemicals: Project cluster for Implementation of novel Strategies (ASPIS);

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101138475. UK participants in SSbD4CheM project are supported by UKRI. CH participants in SSbD4CheM project receive funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

UK Research and Innovation

Insilica is the proud recipient of an NSF SBIR and NIH SBIR grant.
ToxTrack, Inc is a subsidiary of Insilica.

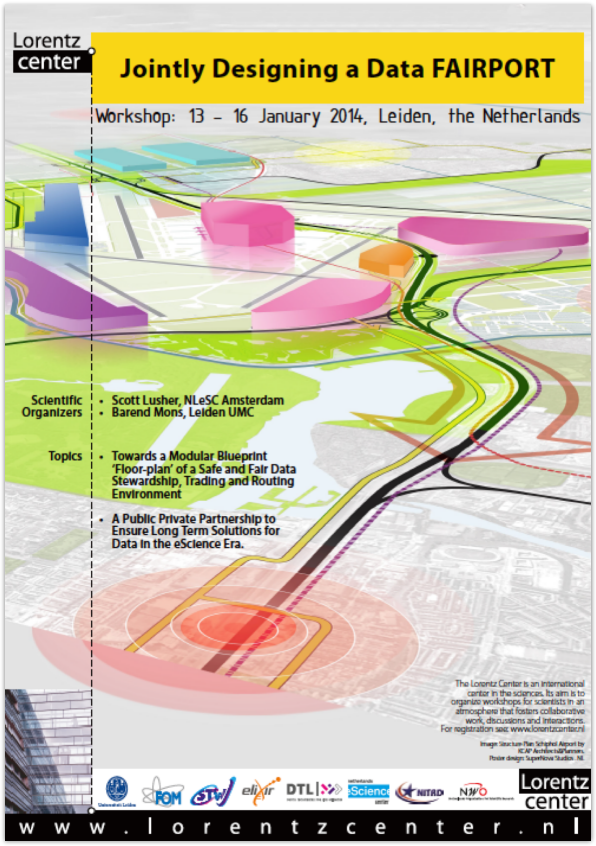
Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

FAIR Principles & AI?

Key Questions:

1. What are the essential elements of data curation and formatting for AI readiness? (What type of data curation and formatting is necessary to ensure that AI is fed with all relevant data?)
2. How can we ensure that all relevant data, including minor parameters, are accurately captured and utilized?
3. Which standards or ontologies should be adopted to improve data interoperability and machine readability?
4. How AI could potentially be used to promote FAIR principles in chemical safety assessment?
5. What measures can be taken to identify and mitigate biases in AI models resulting from data selection or curation?
6. How can we manage the lifecycle of AI models, including updates and decommissioning, in response to changes in data?

https://www.ecetoc.org/wp-content/uploads/2024/10/AI-WS_Programme_FINAL_web.pdf

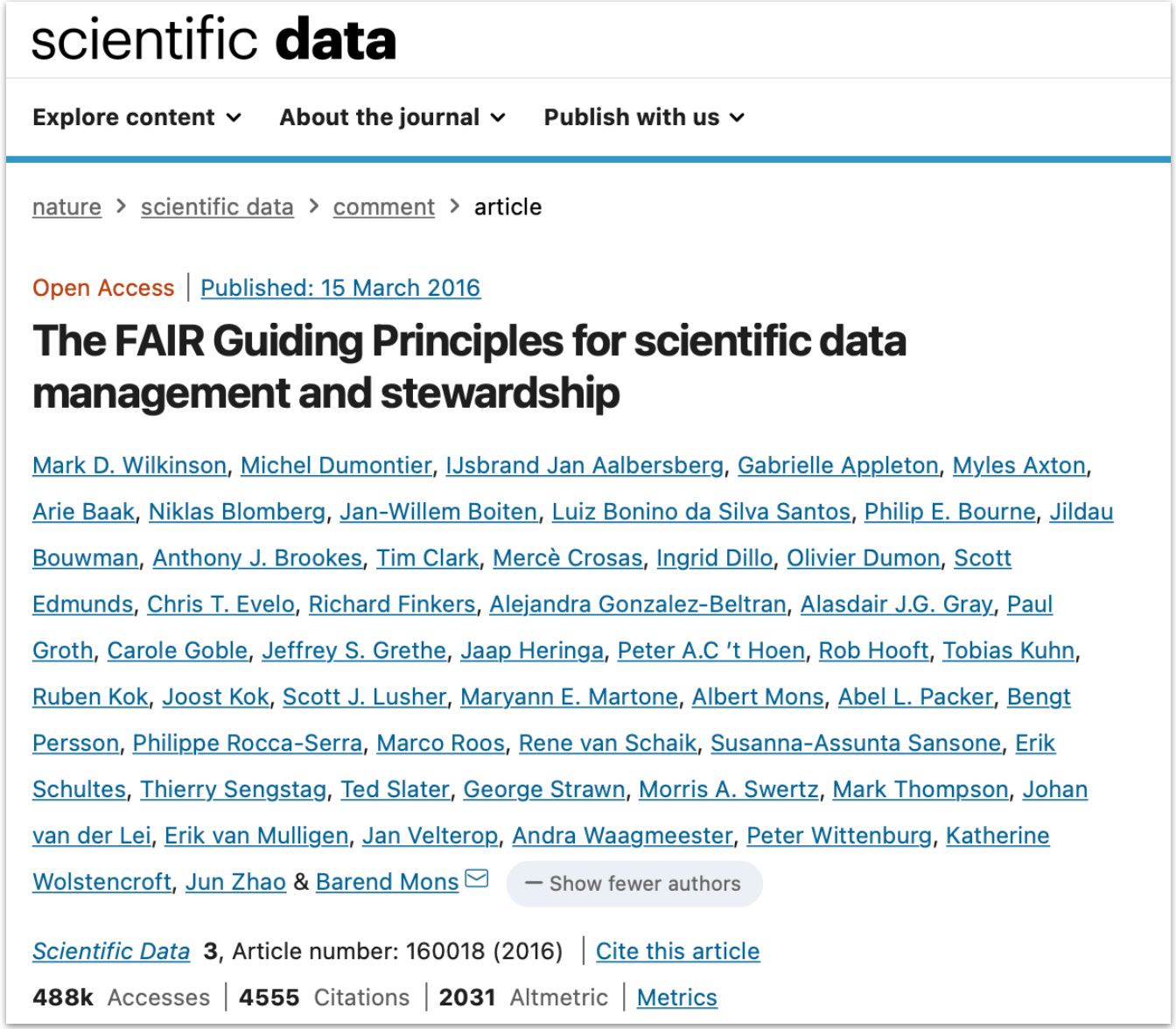

Lorentz center
Jointly Designing a Data FAIRPORT
Workshop: 13 – 16 January 2014, Leiden, the Netherlands
Scientific Organizers
• Scott Lusher, NLeSC Amsterdam
• Barend Mons, Leiden UMC
Topics
• Towards a Modular Blueprint 'Floor-plan' of a Safe and Fair Data Stewardship, Trading and Routing Environment
• A Public Private Partnership to Ensure Long Term Solutions for Data in the eScience Era.
The Lorentz Center is an international center in the sciences. Its aim is to organize workshops for scientists in an atmosphere that fosters collaborative work, discussions and interactions.
For registration see: www.lorentzcenter.nl
Lorentz center
www.lorentzcenter.nl

2014


NIAS
Lorentz center
Workshop @Oort
The Road to FAIR and Equitable Science
22 - 26 January 2024, Leiden, the Netherlands
Scientific Organizers
• Barend Mons, LUMC / LACDR
• Erik Schultes, GO FAIR Foundation / LACDR
• Francisca Oladipo, Thomas Adewumi University
Topics
• Reflection and Looking Ahead
• Machine Actionability
• Equitability
• Fully AI Ready
• Looking Forward to the Next 10 Years
The Lorentz Center organizes international workshops for researchers in all scientific disciplines. Its aim is to create an atmosphere that fosters collaborative work, discussions and interactions.
For registration see: www.lorentzcenter.nl
This workshop is part of our collaboration with NIAS and aims to stimulate research in the humanities & social sciences.
Lorentz center
www.lorentzcenter.nl

2024

2016



scientific **data**

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › scientific data › comment › article

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons — Show fewer authors

Scientific Data **3**, Article number: 160018 (2016) | Cite this article

…the FAIR Principles put specific emphasis on **enhancing the ability of machines to automatically find and use the data**, in addition to supporting its reuse by individuals.

2016

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

2016

Box 2 | The FAIR Guiding Principles

Computability

- *FAIR data* use knowledge representation languages and controlled vocabularies that reduce/eliminate ambiguity.

Trustworthiness

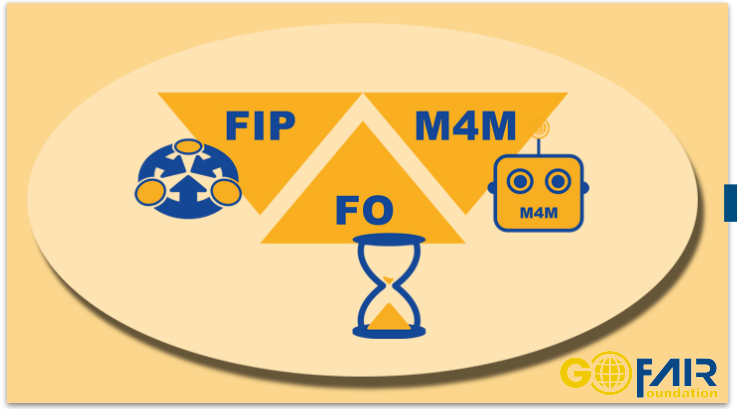
- *FAIR data* have (rich) provenance which provides evidence for the source of data. Provenance can include uncertainties and error estimation. Trust also requires large amounts of metadata.

Equitabilty

- *FAIR data* make explicit the conditions for reuse. Equitabilty can be technologically ensured, when data and metadata are FAIR

First decade of FAIRification…

FAIR Practices & Training

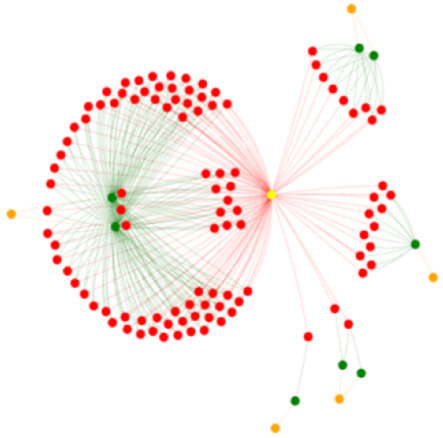


CEDAR Workbench

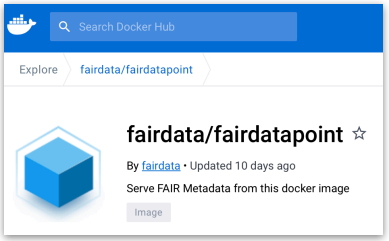
FIP Wizard

FAIR Connect

FAIR Knowledge Graphs



FAIR Orchestration



NANODASH

CEDAR Workbench

SRDC

REDCap
Research Electronic Data Capture

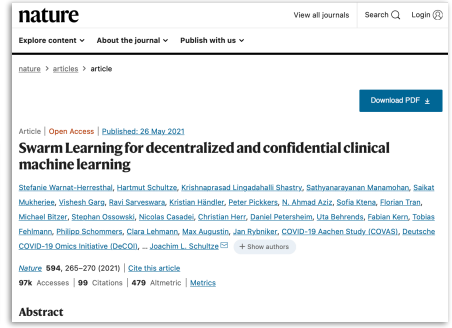
SeSaMe
SemanticSerchMetadata

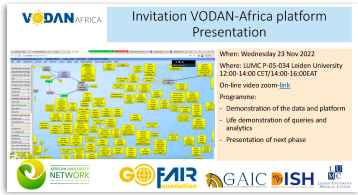
MOLGENIS

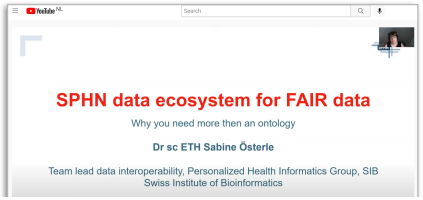
castor

FAIR Solutions

DESCRIBO
DESCRIBE EVERYTHING







Data Visiting

Integrating AI into chemical safety assessment
Opportunities, challenges, and the path forward

Challenges

Key Considerations:

- **Filtering:** The data collected from these sources are heavily filtered to remove noise, low-quality text, and duplicates.

- **Bias and Ethical Concerns:** Training on public data can introduce biases present in the original content (e.g., social biases, regional biases), which is why researchers try to apply techniques to mitigate these issues.

- **Copyrighted Material:** While efforts are made to use publicly available content, there are ongoing debates and concerns about training on data that may come from copyrighted works, especially when it is scraped without explicit permission.

Content generated using OpenAI's ChatGPT-4

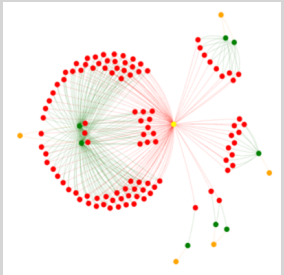
FAIR Considerations:

Key Considerations:

Computability — **Filtering:** The data collected from these sources are heavily filtered to remove noise, low-quality text, and duplicates.

Trustworthiness — **Bias and Ethical Concerns:** Training on public data can introduce biases present in the original content (e.g., social biases, regional biases), which is why researchers try to apply techniques to mitigate these issues.

Equitability — **Copyrighted Material:** While efforts are made to use publicly available content, there are ongoing debates and concerns about training on data that may come from copyrighted works, especially when it is scraped without explicit permission.

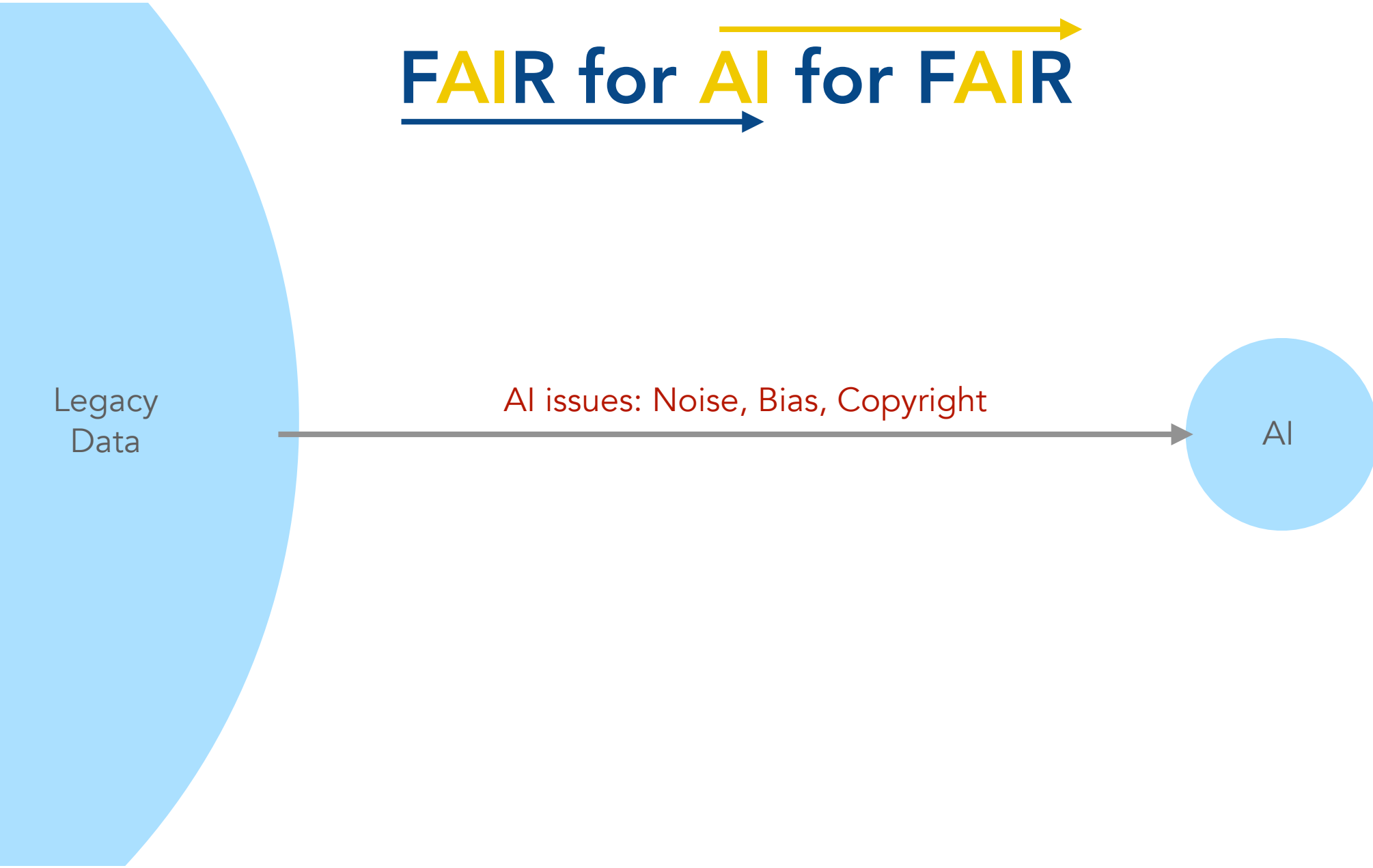


Content generated using OpenAI's ChatGPT-4

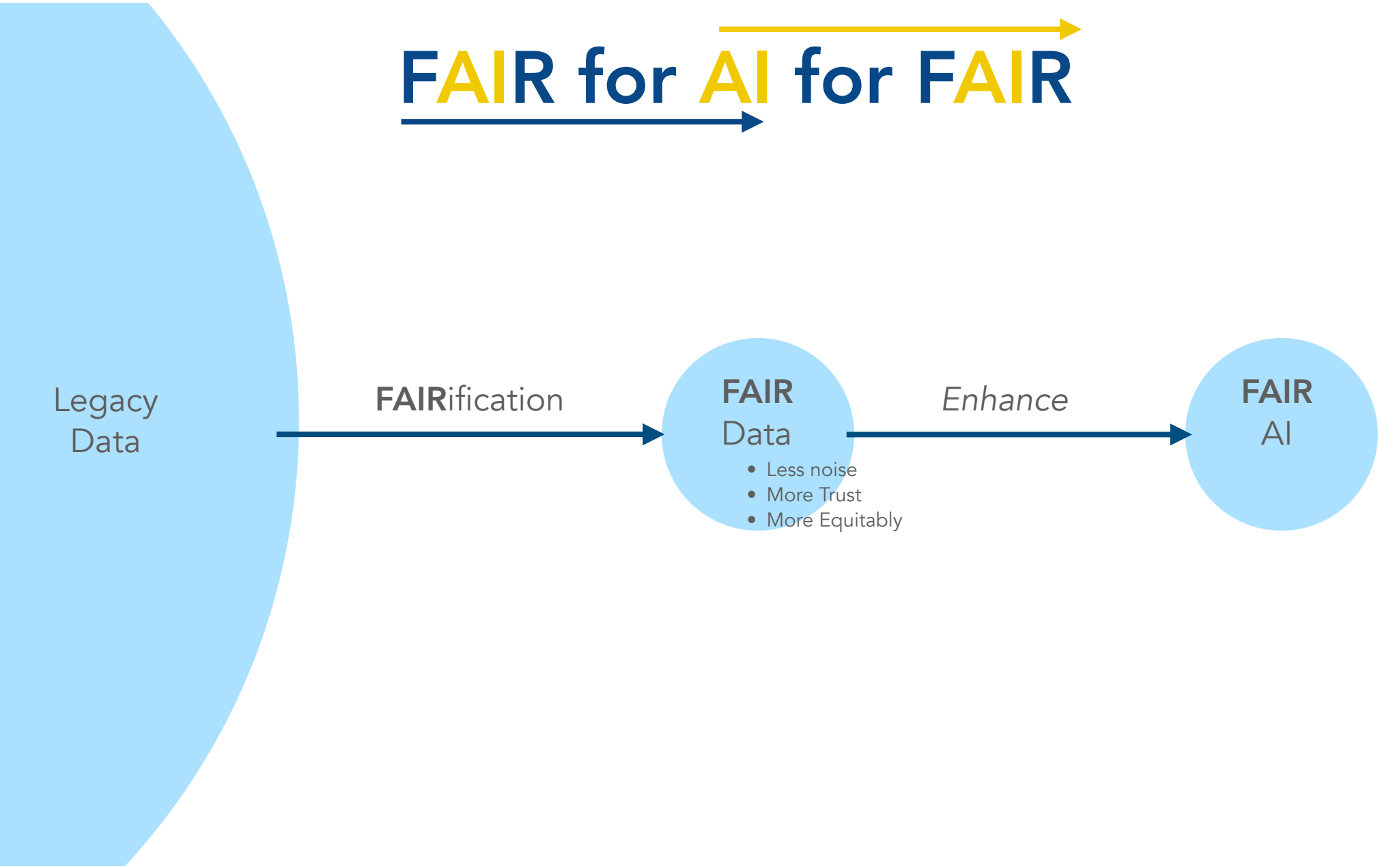
Integrating AI into chemical safety assessment
Opportunities, challenges, and the path forward

Path forward…

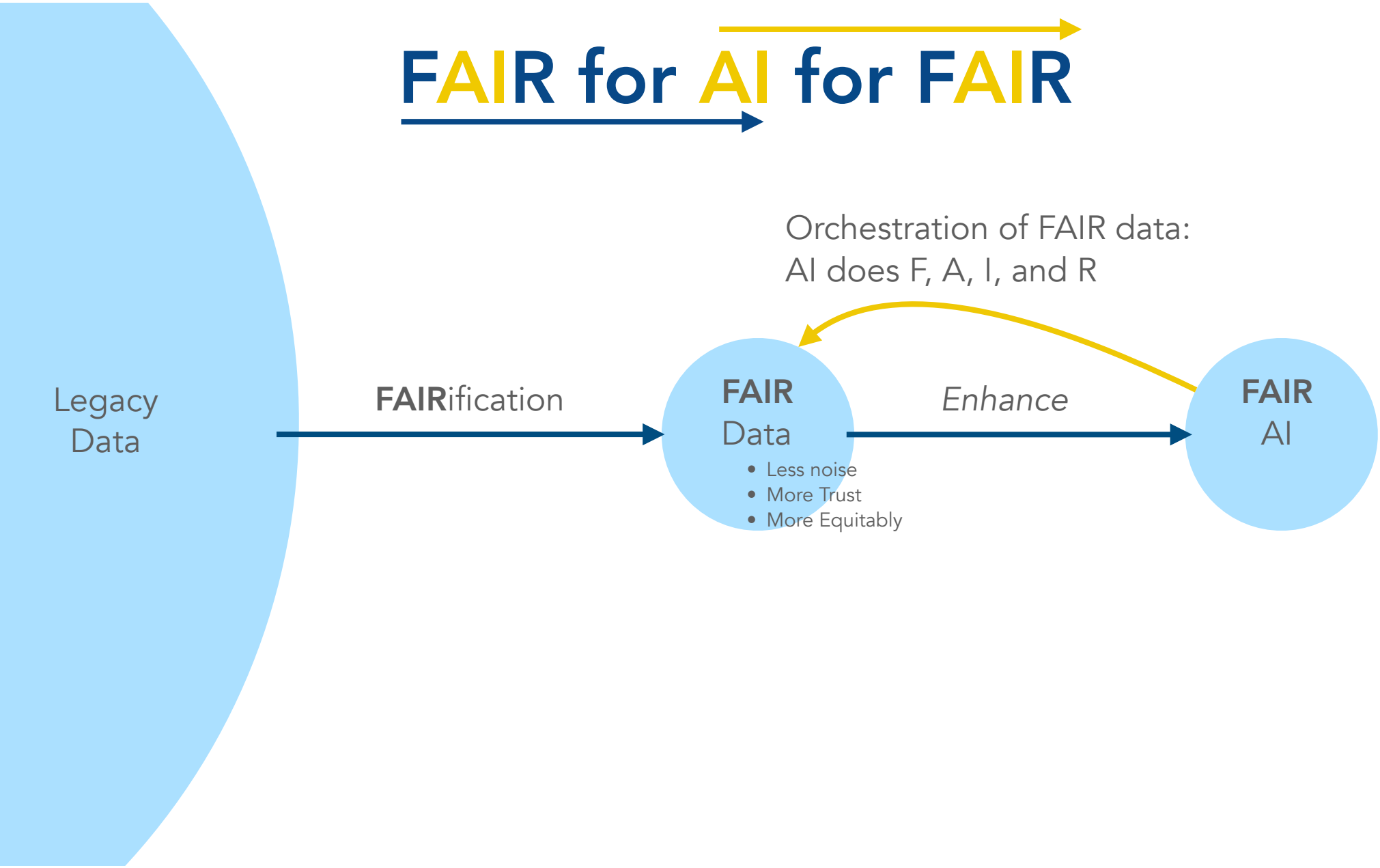
FAIR for AI for FAIR



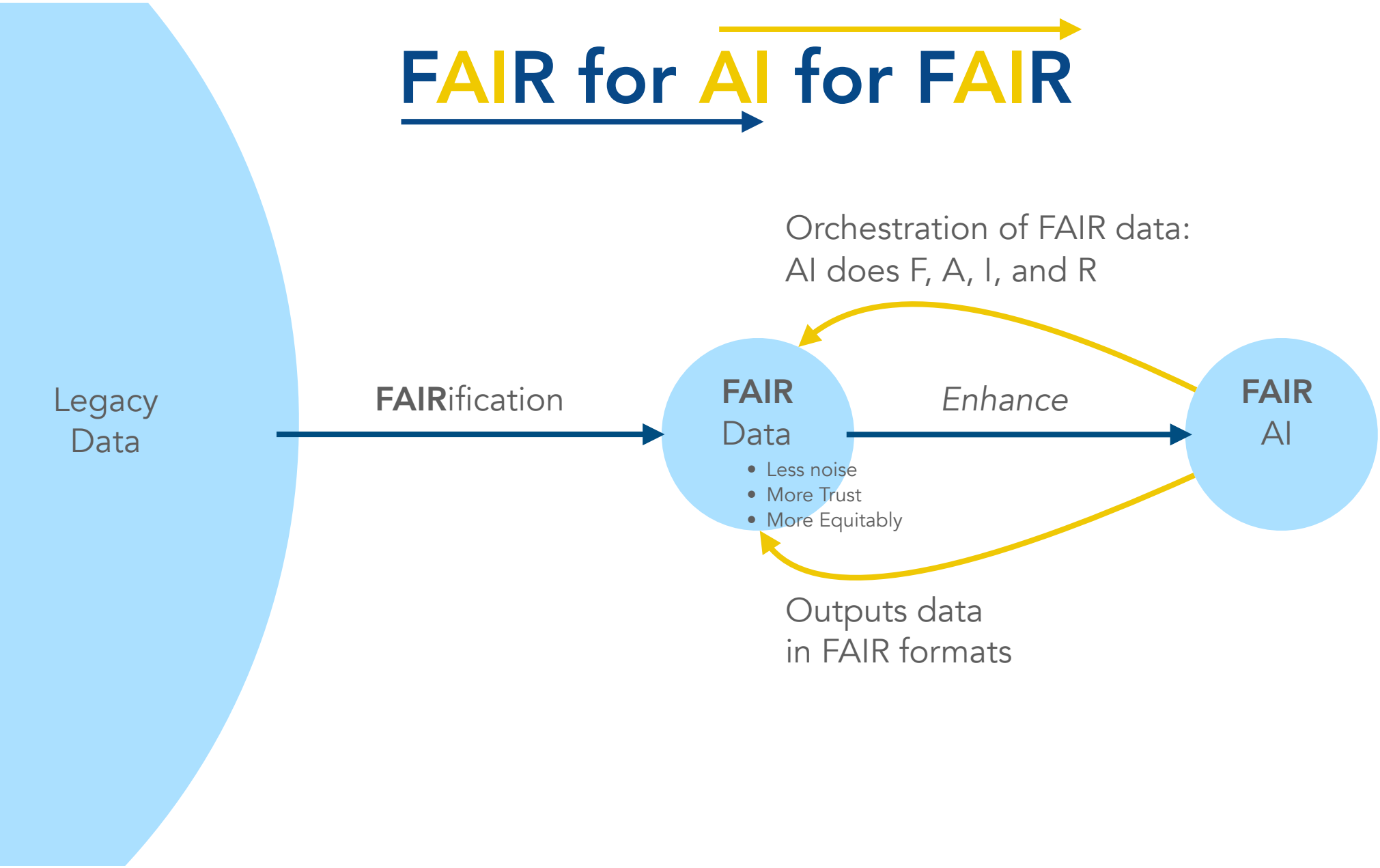
FAIR for AI for FAIR



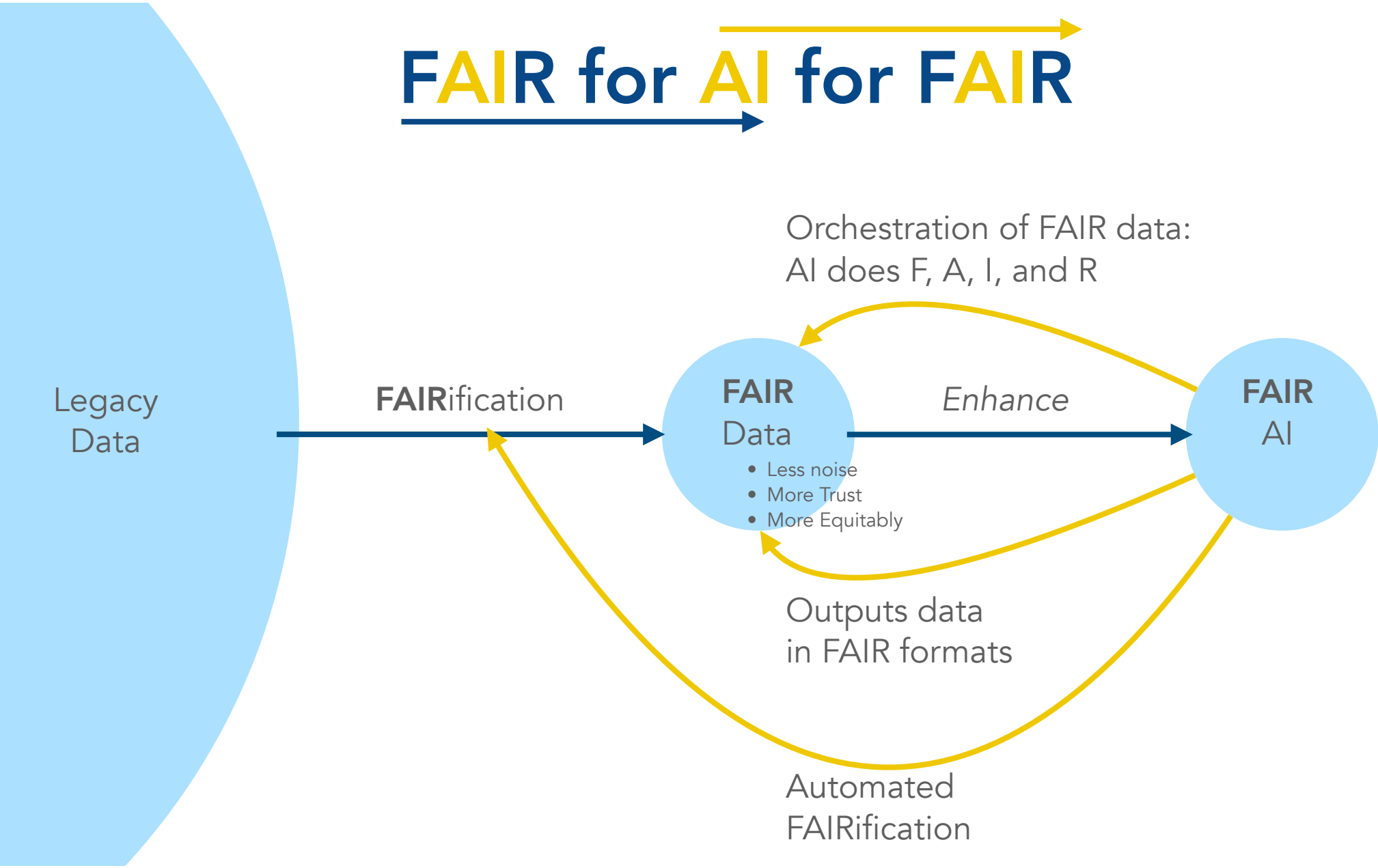
FAIR for AI for FAIR

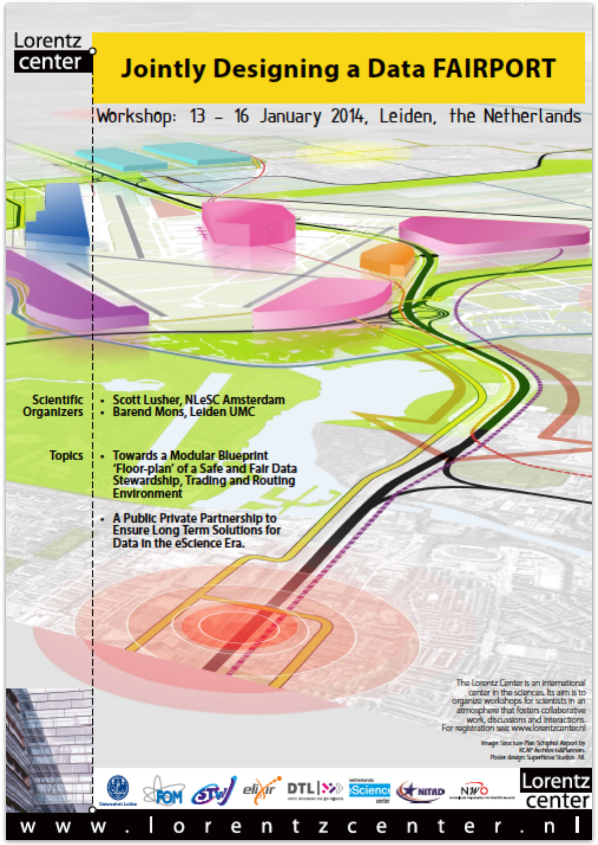


FAIR for AI for FAIR



FAIR for AI for FAIR





2014

FAIR as in "fair"

2024

FAIR as in "AI-Ready"

See also: FAIR in ML, AI Readiness, & Reproducibility (FARR) Workshop October 9-10 https://www.farr-rcn.org/workshop24

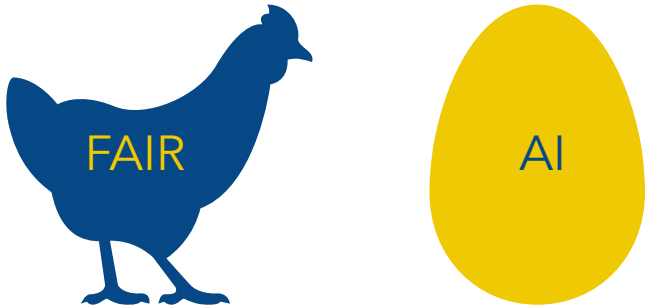
OpenAI
Microsoft
AI
Amazon Q
Google AI

Integrating AI into chemical safety assessment
Opportunities, challenges, and the path forward

ecetoc

October 16 2024
https://www.ecetoc.org/event/ai-workshop/

FAIR well



Erik Schultes
http://orcid.org/0000-0001-8888-635X
eriks@gofair.foundation
These slides: https://osf.io/ueqvw

GO FAIR foundation

LACDR

LIFES Leiden Institute for FAIR & Equitable Science