



The PhUSE Basel – SDE
Future of Analytics

A Framework for Making Data FAIR at Bayer



Future of Analytics

June 2018 / Alexandra Grebe de Barron
& Drashti Vasant





Agenda

- // Where we stand today
- // The guiding principles: FAIR
- // Implementing the FAIR Framework @Bayer



Where do we stand today?

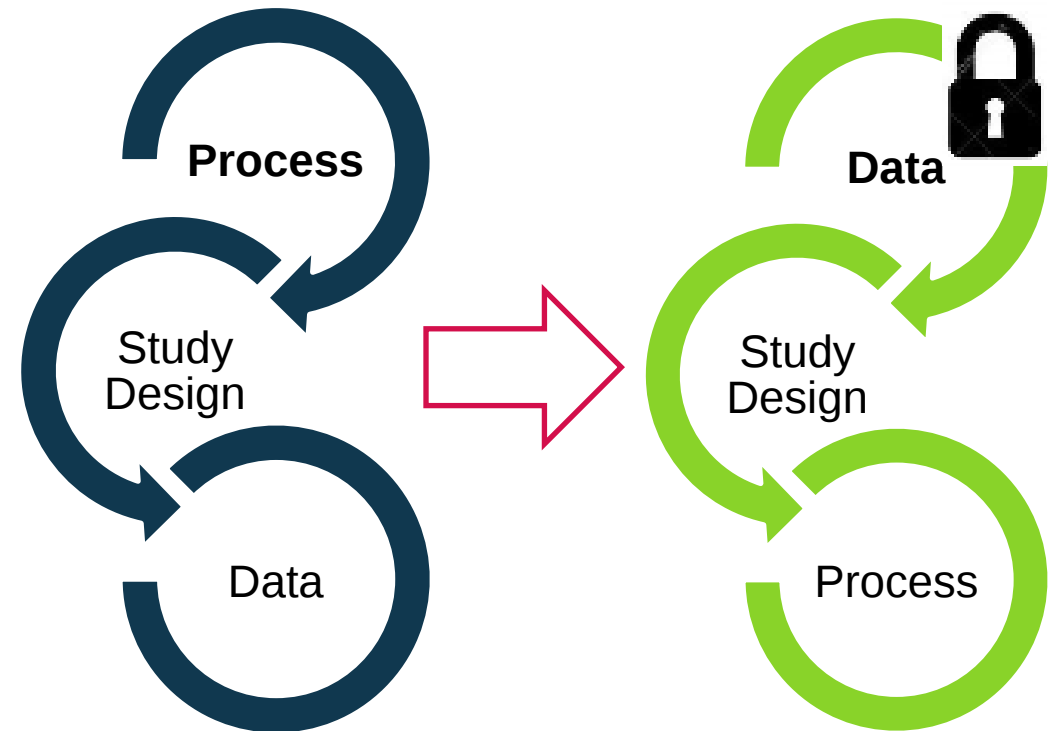
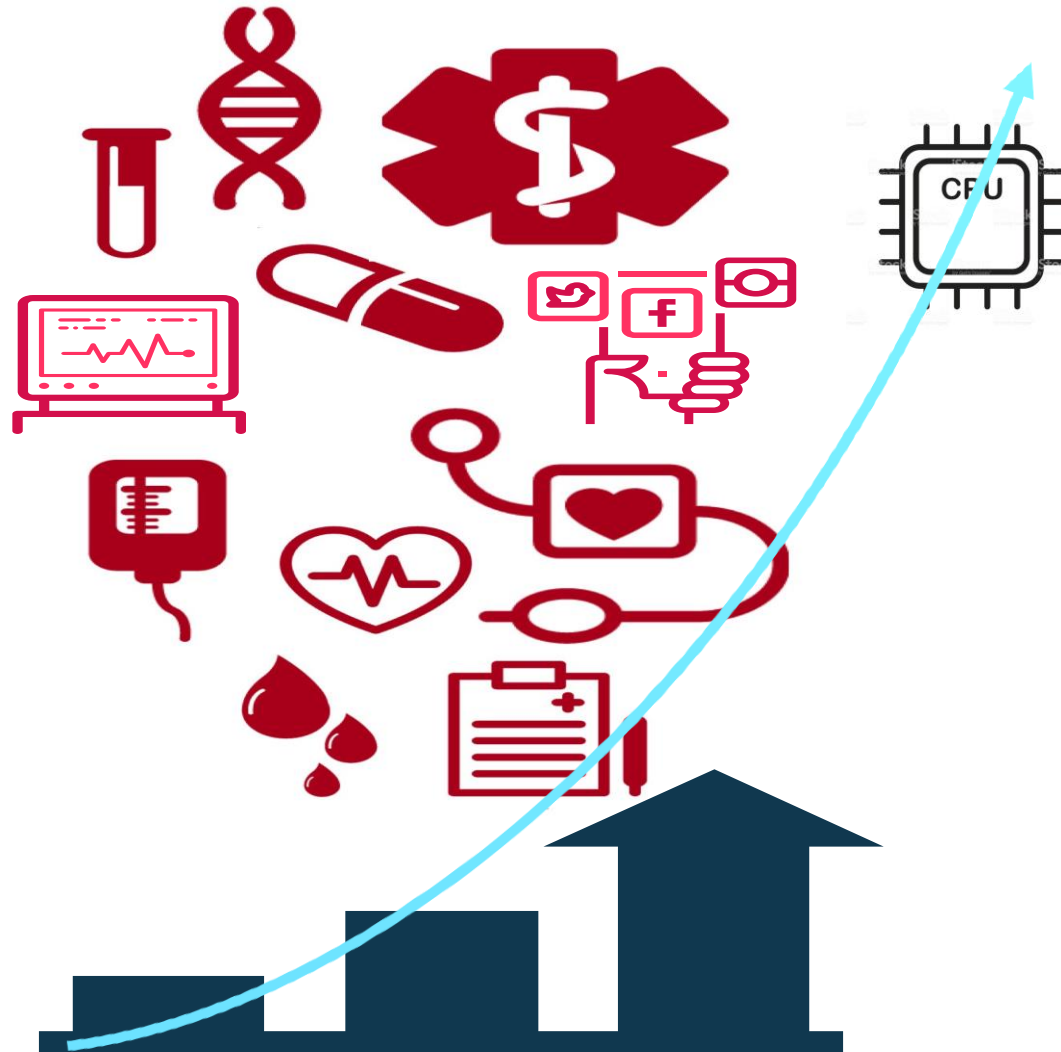
Changing Data landscape

Digital Agenda

AI



Change in strategy...



The guiding principles: FAIR data



*F*indable *A*ccessible *I*nteroperable *R*eusable



FAIR DATA – What is it?



- // 'FAIR Guiding Principles for scientific data management and stewardship' were published in *Scientific Data*, 2016
- // Developed by a **consortium of stakeholders** from academia, industry, funding agencies, and scholarly publishers
- // They designed and jointly endorsed a **concise and measureable set of principles**
- // They put specific emphasis on enhancing the ability of **machines** to **automatically find and use the data**

Source: <https://www.force11.org/group/fairgroup/fairprinciples>



Fair data principles – support infrastructure

Findable

- F1. (meta)data are assigned a globally unique and persistent identifier;**
- F2. data are described with rich metadata;**
- F3. metadata clearly and explicitly include the identifier of the data it describes;**
- F4. (meta)data are registered or indexed in a searchable resource;**

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;**

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;**

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;**

Myth bursters





FAIR Framework within Bayer

// People:

Set up FAIR data stewardship teams
(roles, responsibilities, trainings)



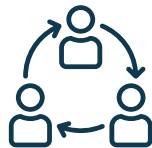
// Processes:

Set up FAIRification process
(FAIR assessment, FAIR actions)



// Regulations:

Set up operational instructions,
governance bodies, legal framework



// Systems:

Set up FAIR data infrastructure
(best practices, IT landscape
recommendations, solutions)





Implementing the Framework @Bayer

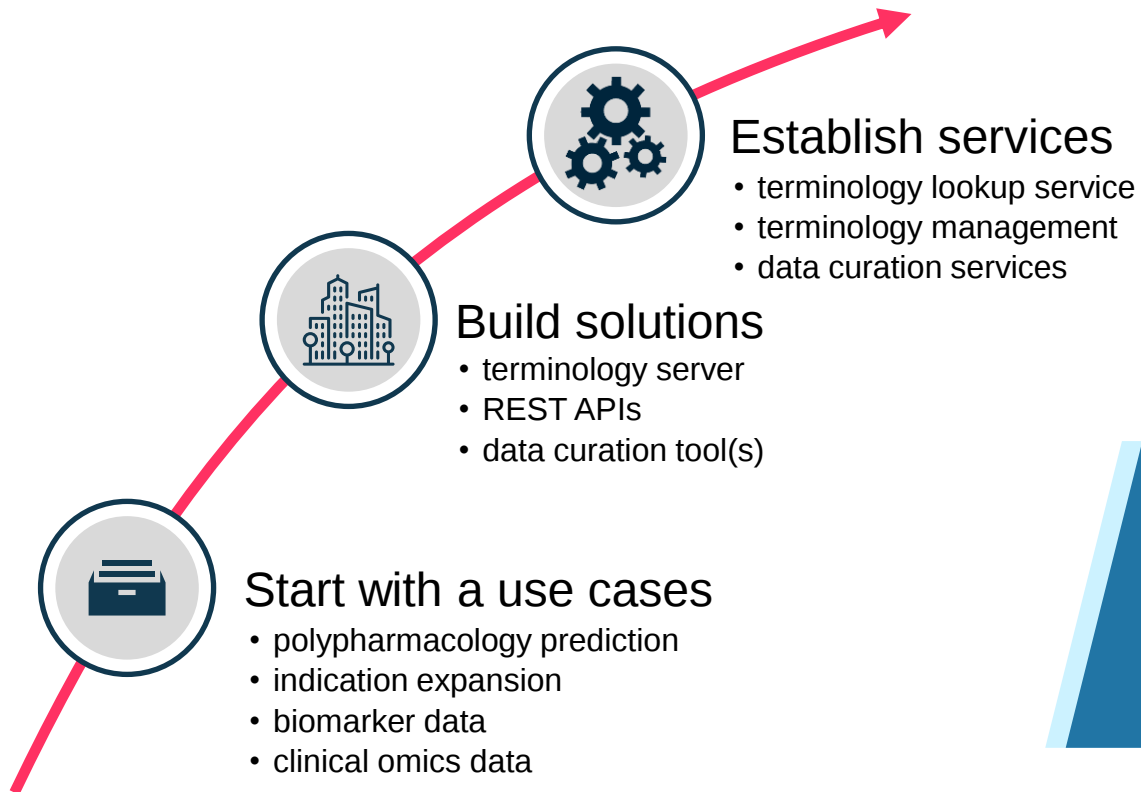
bottom up – use case driven

top down – strategic program

**collaborative – cross pharma
community**

Bottom up approach for data integration

Use case driven implementation project



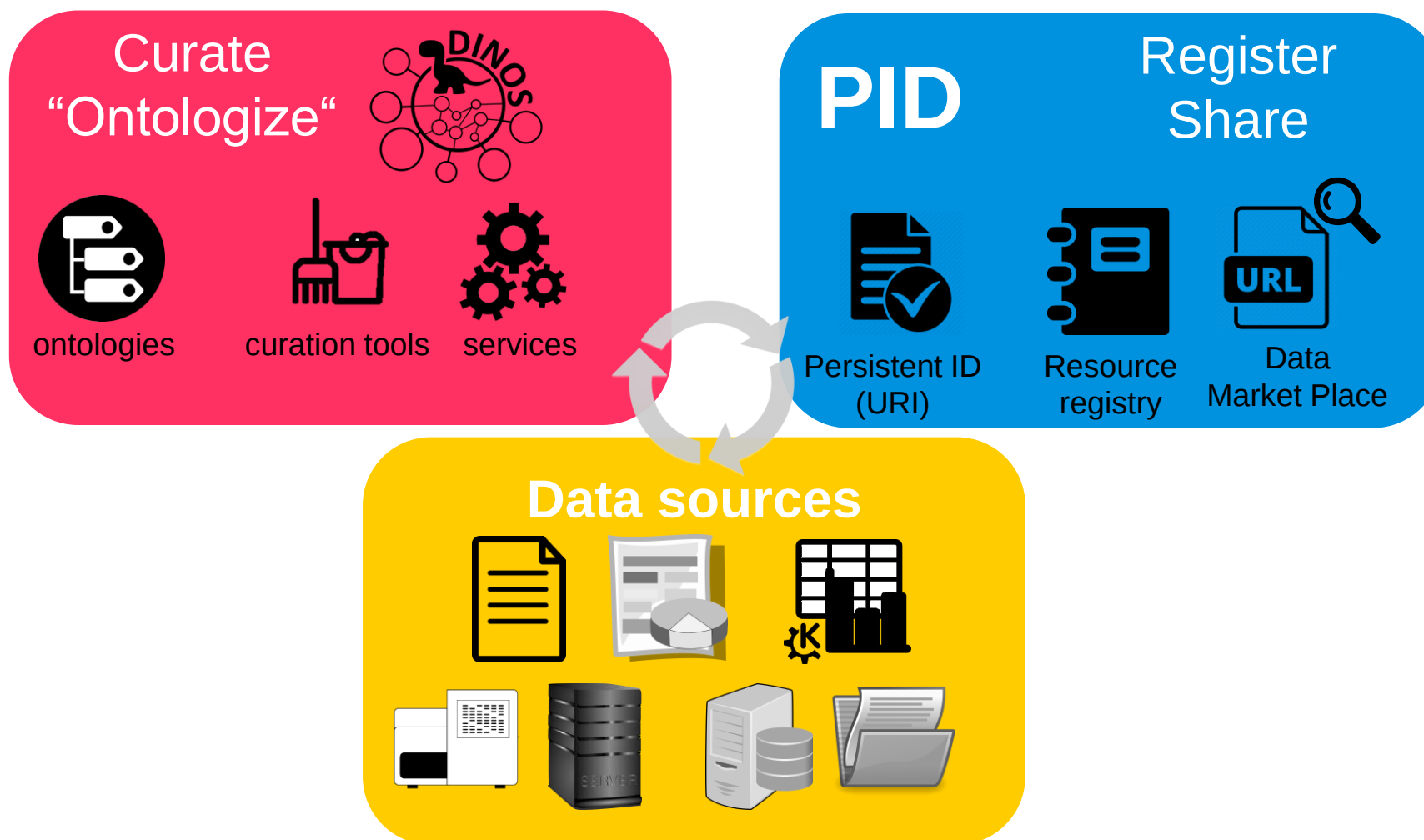
Common data scientist problem:

- // Data not analysis-ready
- // No common controlled vocabulary
- // No sustainable curation process



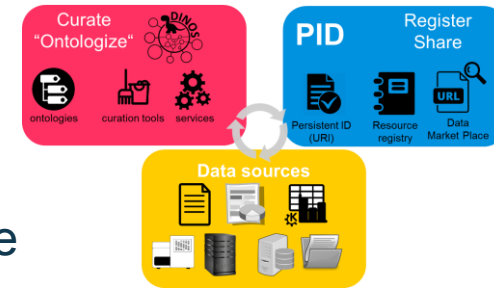
Bottom up approach for data integration

Building the basic infrastructure for making data FAIR



Make & keep data FAIR

We implement the FAIR data principles as promoted by the „Data as an asset“ initiative



To be Findable

- // use unique and persistent identifiers
- // register in a searchable resource
- // **PID and Ontology lookup service**

To be accessible

- // data are retrievable by their identifier using a standardized communications protocol
- // **REST APIs, SPARQL endpoint**

FAIR data principles

To be interoperable

- // use a broadly applicable language for knowledge representation
- // use vocabularies that follow FAIR principles
- // **Ontology based curation, ontology mapping and use of RDF/OWL**

To be re-usable

- // data have accurate, relevant attributes
- // data are associated with their provenance.
- // data meet domain-relevant community standards
- // **Use public ontologies where applicable, highly curated datasets and minimum meta data model**

Top-Down Approach: The Digital excellence council to ensure a group-wide digital transformation approach



Digital embedded in strategies

// Pharmaceuticals

// Consumer Health

// Crop Science

// Corporate Functions/Business Services



Strategic commitment on FAIR data / budget



Organisational changes / new roles / trainings



Mindset change towards data sharing / awareness

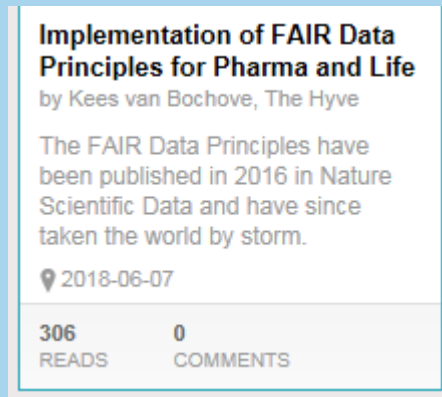


Implementation



Collaboration in cross-pharma / academic communities

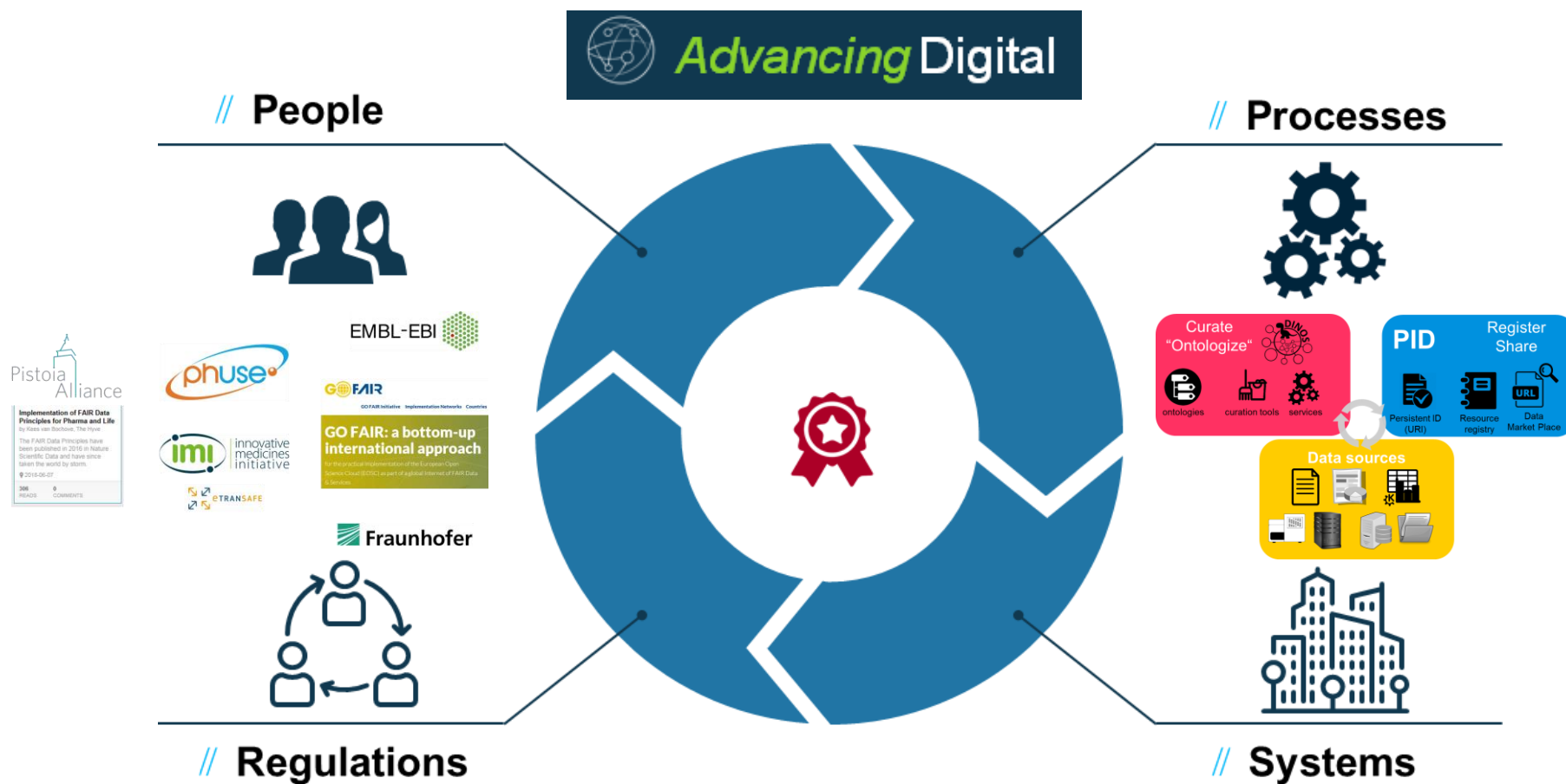
Principles – Guidance – Best Practices - Synergies



<https://ip3.pistoiaalliance.org/>



Implementing the Framework @Bayer



- // FAIR-ification of data according GO-FAIR
- // Hackathons
- // FAIR assessment
- // Mindset change
- // Data as an asset
- // Data sharing incentives
- // Organisational
- // FAIR data training
- // FAIR data teams



Thank you!

////////

For more information:

drashti.vasant@bayer.com

alexandra.grebedebarron@bayer.com





Forward-Looking Statements

This website/release/presentation may contain forward-looking statements based on current assumptions and forecasts made by Bayer management.

Various known and unknown risks, uncertainties and other factors could lead to material differences between the actual future results, financial situation, development or performance of the company and the estimates given here. These factors include those discussed in Bayer's public reports which are available on the Bayer website at <http://www.bayer.com/>.

The company assumes no liability whatsoever to update these forward-looking statements or to conform them to future events or developments.



Abstract

We are in the age of digital economy where data is considered as the oil of the 21st Century (1). As the pharmaceutical industry makes a strategic decision to shift from molecule- to data-driven business models, there is an urgent need to start handling our data smartly. We, at Bayer have also intensified our work towards transforming our data into knowledge.

In order to make data both machine and human readable, so as to enable downstream processes such as advanced analytics, machine learning and artificial intelligence, we have set out to create a *centralized infrastructure* to support curation and integration of our data.

Guiding us to make the right decision towards building the correct framework in-house are the FAIR principles. We are working towards:

- // Systematic assessment (fit/gap analysis) of data in regard to the FAIR data principles resulting in a maturity classification of our data concerning FAIR
- // Follow up actions to raise the maturity level of our data concerning one or all FAIR data principles depending on the demand.

And in parallel have set up

- // Central services such as ontology management, data curation services, building canonical data models.

Our methods in building this framework are collaborative and use-case driven:

- // We aim for collaboration with the FAIR data community to set up Pharma community standards for the FAIR data assessment and follow up actions.
- // We plan the actions around FAIR data by focusing on use cases that create business value. This helps us to scope our activities to the right level of detail.

We will present our first results on DINOS (Data INtegration and Ontology Services), our foundation for the FAIR data framework.