"FAIR-by-Design" Artifacts:
Enriching Publications and Software with FAIR Scientific Information at the Time of Creation

Oliver Karras, Patrick Kuckertz, Jan Göpfert, Tristan Pelser, Rodrigo Pueblas, Jann M. Weinand, Detlef Stolten, Sören Auer
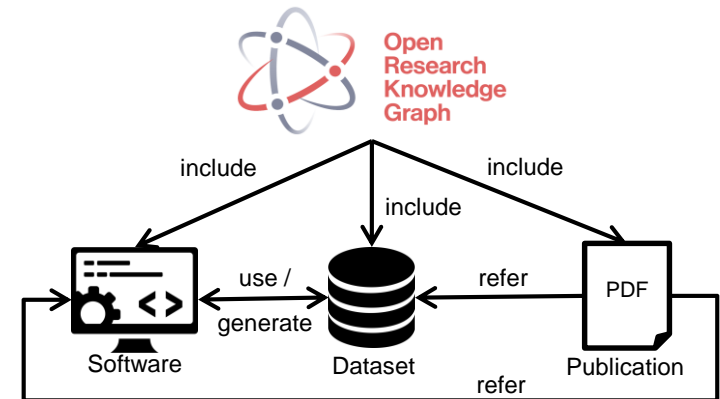
Contact: oliver.karras@tib.eu

28 September 2023, NFDI4Ing Conference

TIB LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

JÜLICH
Forschungszentrum

Introduction

Publications, Software, and Datasets are Core Artifacts in Academia

- Need for **infrastructures**, **services**, and **tools** to organize FAIR scientific information from all of three artifact types

- Ellen uses Open Research Knowledge Graph (ORKG)
  - Cross-discipline RKG including **all three artifact types**
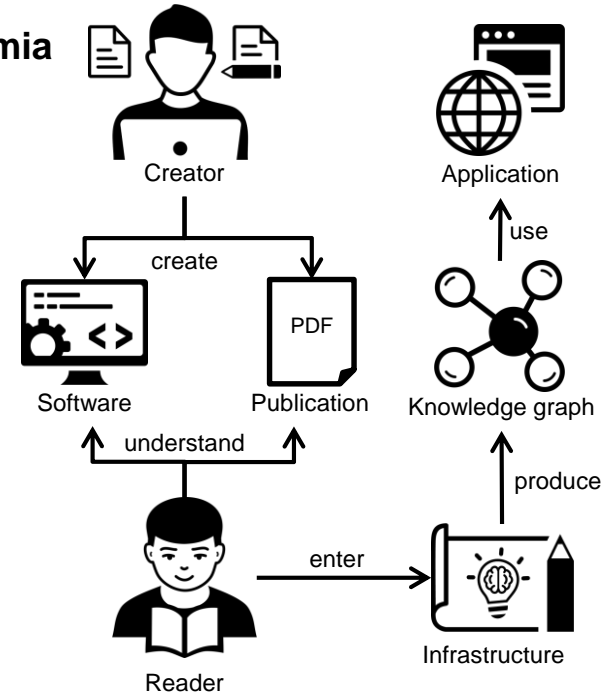 - Use of **crowd-sourcing** and **(semi-)automated approaches** for organizing scientific information

Introduction

Publications, Software, and Datasets are Core Artifacts in Academia

- Need for **infrastructures**, **services**, and **tools** to organize FAIR scientific information from all of three artifact types

- Ellen uses <u>Open Research Knowledge Graph</u> (ORKG)
  - Cross-discipline RKG including **all three artifact types**
  - Use of **crowd-sourcing** and **(semi-)automated approaches** for organizing scientific information
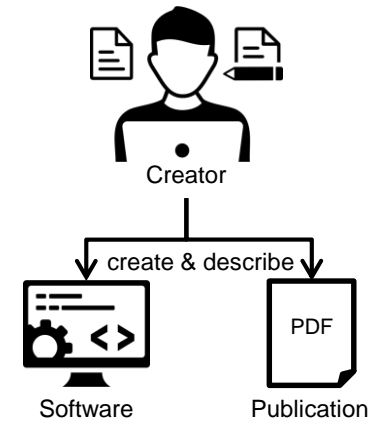
Problem:
ORKG focuses on **published** artifacts, so **FAIRification** is a **downstream task** and often **not done by the creators** themselves.

Creator — create → Software, Publication

Application — use

Knowledge graph — produce

Reader — understand (Software, Publication)

Reader — enter → Infrastructure

Ellen's Idea: "FAIR-by-Design" Artifacts
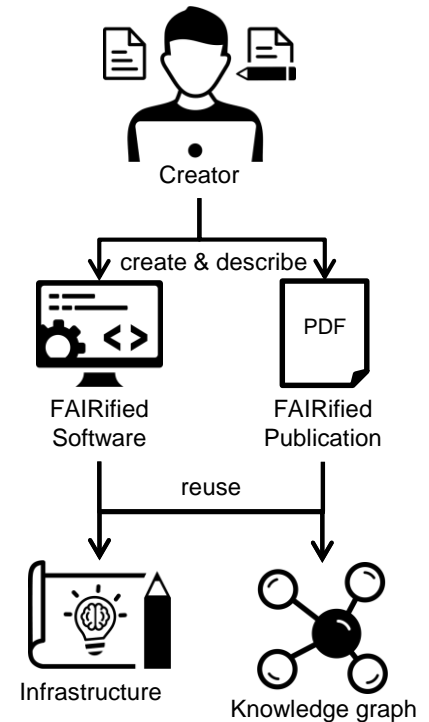
FAIRification of Artifacts at the Time of their Creation

●— **Creators** of artifacts **describe** them with FAIR information **only once** and **in parallel** at the time of **creation**

Creator

create & describe

Software   Publication

Ellen's Idea: "FAIR-by-Design" Artifacts

FAIRification of Artifacts at the Time of their Creation

- **Creators** of artifacts **describe** them with FAIR information **only once** and **in parallel** at the time of **creation**

- **Embedded** FAIR information **into** the **artifact** itself
  - Information **persists for the lifetime** of the artifact
  - Information is **available for anyone at any time**
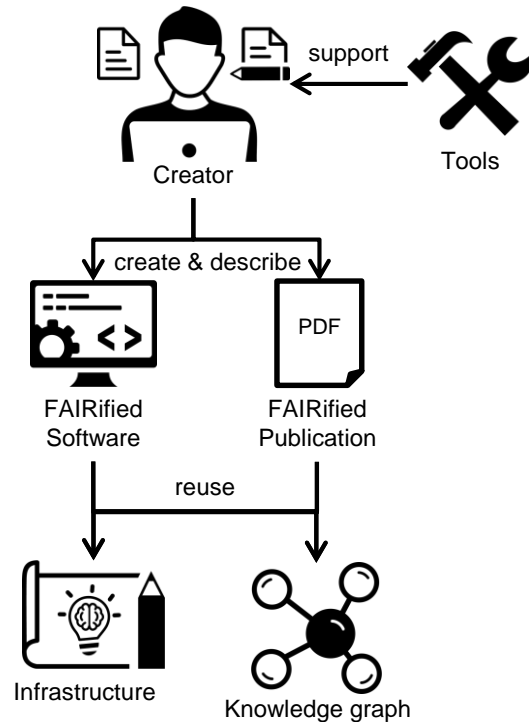 - **Reuse** information, e.g., import into RKGs

Creator

create & describe

FAIRified Software

PDF

FAIRified Publication

reuse

Infrastructure

Knowledge graph

Ellen's Idea: "FAIR-by-Design" Artifacts

FAIRification of Artifacts at the Time of their Creation

- **Creators** of artifacts **describe** them with FAIR information **only once** and **in parallel** at the time of **creation**

- **Embedded** FAIR information **into** the **artifact** itself
  - Information **persists for the lifetime** of the artifact
  - Information is **available for anyone at any time**
 - **Reuse** information, e.g., import into RKGs

> **Idea:**
> Develop **tools** that support creators in
> **FAIRifying** artifacts at the time of their **creation**.

support

Creator

Tools

create & describe

PDF

FAIRified
Software

FAIRified
Publication

reuse

Infrastructure

Knowledge graph

NFDi4ing

Ellen's Tools for "FAIR-by-Design" Publications & Software

ORKG, SciKGTeX, and DataDesc along the RDM Lifecycle
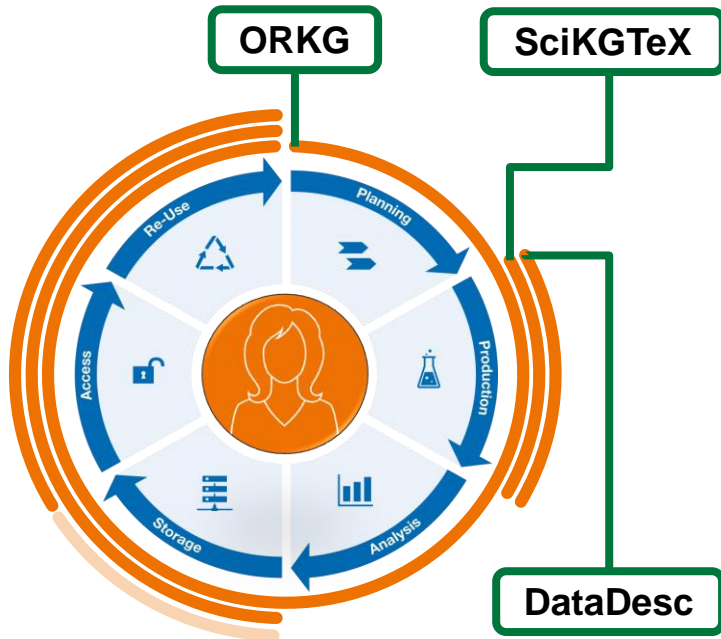
SciKGTeX – For publications
- Authors can **annotate** research contributions directly in the **LaTeX source code**
- SciKGTeX **embeds** the structured and machine-actionable research contributions **into** the **PDF's XMP metadata**
- ORKG **supports import** of SciKGTeX annotations

DataDesc – For software
- Developers can **annotate** individual API functions and parameters directly within the **source code**
- DataDesc **converts** all metadata **into** an OpenAPI-compliant **YAML file**
- ORKG **import** is **ongoing (proof of concept)**

SciKGTeX – Scientific Knowledge Graph TeX

FAIRification of Publications While Writing

- LaTeX commands to annotate **main properties** of a scientific contribution
  - **5 predefined commands**: Research problem, objective, method, result, and conclusion
  - Support for own **custom properties**
  - Documentation: https://github.com/Christof93/SciKGTeX

Full paper with all details and evalution:
Bless et al.: SciKGTeX – *A LaTeX Package to Semantically Annotate Contributions in Scientific Publications*, 23nd ACM/IEEE Joint Conference on Digital Libraries, ACM, 2023, **Vannevar Bush Best Paper Award.**

```
\usepackage{scikgtex}
\begin{document}
The role of \researchproblem{antibiotic therapy} is controversial.
The purpose of this study was to \objective{determine the
effectiveness of high-dose amoxicillin/potassium clavulanate in the
treatment of children}.
This was a \method{randomized, double-blind, placebo-controlled
study}.
\result{Children receiving the antibiotic were more likely to be
cured (50% vs 14%) than children receiving the placebo}.
\conclusion{Amoxicillin/potassium clavulanate results in
significantly more cures and fewer failures than placebo}.
\end{document}
```

```
\documentclass{article}
\usepackage{scikgtex}
\addmetaproperty[amo, http://purl.org/spar/amo#]{has_claim}
\begin{document}
We make the claim that \contribution{has_claim}{the earth is round}.
...
```

SciKGTeX – Scientific Knowledge Graph TeX

FAIRification of Publications While Writing – RDF Metadata Example

```xml
<?xpacket begin="?" id="731960eb-9a9c-4996-c9a2-0c296941c6"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/">
<rdf:RDF
  xmlns:orkg="http://orkg.org/core#"
  xmlns:orkg_property="http://orkg.org/property/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <rdf:Description rdf:about="https://www.orkg.org/orkg/paper/731960eb-9a9c-4996-c9a2-0c296941c6">
    <rdf:type rdf:resource="http://orkg.org/core#Paper"/>
    <orkg:hasTitle>SciKGTeX - A  Package to Semantically Annotate Contributions in Scientific Publications</orkg:hasTitle>
    <orkg:hasResearchContribution>
      <orkg:ResearchContribution rdf:about="https://www.orkg.org/orkg/paper/731960eb-9a9c-4996-c9a2-0c296941c6/contribution_ORKG_default">
        <orkg_property:researchproblem>crowd-sourcing for scientific knowledge graphs</orkg_property:researchproblem>
        <orkg_property:objective>workflow for authors of scientific documents to specify their contributions</orkg_property:objective>
        <orkg_property:objective>automatically upload contributions to a knowledge graph</orkg_property:objective>
        <orkg_property:method>latex</orkg_property:method>
        <orkg_property:method>luatex</orkg_property:method>
        <orkg_property:result>score of 79 out of 100 on the System Usability Scale</orkg_property:result>
        <orkg_property:result>7 minutes on average to annotate the main contributions</orkg_property:result>
        <orkg_property:conclusion>SciKGTeX simplifies the process of manual semantic annotation of research contributions in scientific
        articles. Our workflow demonstrates how a scientific knowledge graph can automatically ingest research contributions from document
        metadata.</orkg_property:conclusion>
      </orkg:ResearchContribution>
    </orkg:hasResearchContribution>
  </rdf:Description>
</rdf:RDF>
</x:xmpmeta>
<?xpacket end="r"?>
```
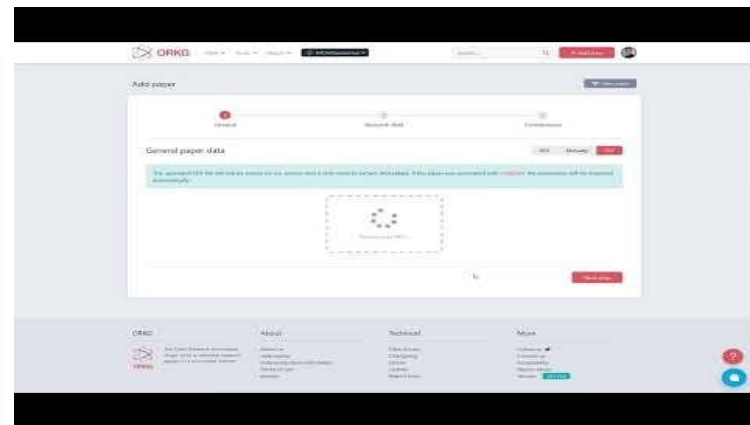
SciKGTeX – Scientific Knowledge Graph TeX

FAIRification of Publications While Writing – Import into ORKG

https://www.youtube.com/watch?v=ZzrQ_YCKVsYa

NFDi4ing

DataDesc

FAIRification of Software While Programming

- **Metadata schema** for software documentation focusing on describing **interfaces** compliant to
  - OpenAPI
  - Schema.org
  - CodeMeta
- Machine-actionable **metadata exchange format**
- Software **toolkit** supporting **documentation**, **extraction**, and **publication** of software metadata

> **Preprint with all details and application case:**
> Kuckertz et al.: *A Metadata-Based Ecosystem to Improve the FAIRness of Research Software*, arXiv preprint arXiv:2306.10620, 2023.

API Function Object

Property	Label	Data Type	Required
identifier	Identifier	Text	True
deprecated	This object is deprecated	Boolean	False
inputVariables	Input variables	Variable Object(s)	False
outputVariables	Output variables	Variable Object(s)	False

Variable Object

Property	Label	Data Type	Required
identifier	Identifier	Text	True
required	Value input is required	Boolean	False
deprecated	This object is deprecated	Boolean	False
dataSchema	Data schema	Data Schema Object	True

Data Schema Object

Property	Label	Data Type	Required
identifier	Identifier	Text	False
semanticConcept	Semantic concept reference	Text or URI	False
type	Data type	Text or Data Schema Object(s)	True
format	Data format	Text	False
minimum	Minimum value	Number	False
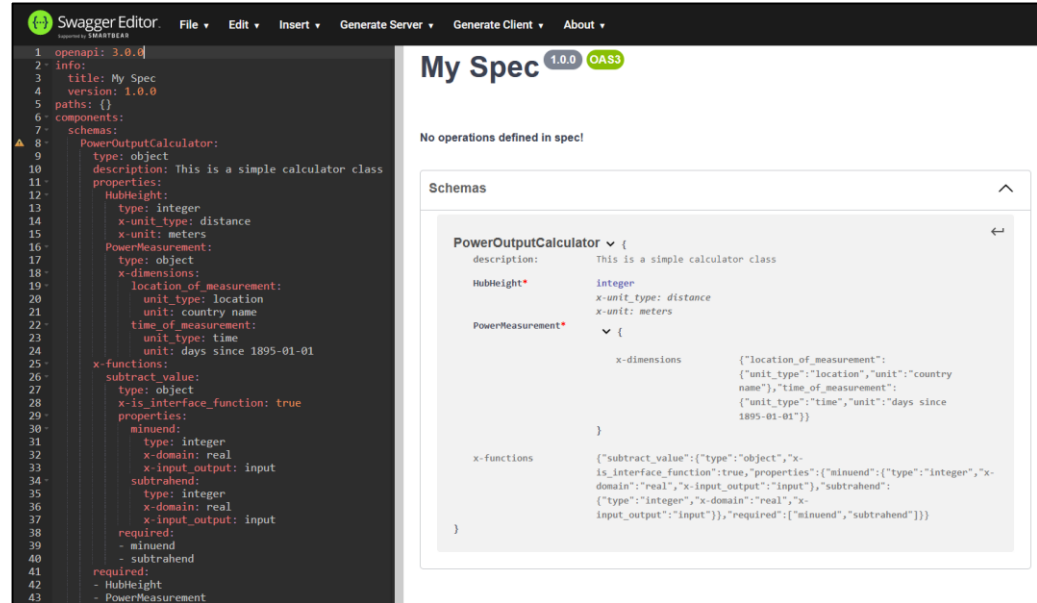
DataDesc

FAIRification of Software While Programming – Workflow

Interface annotation
in source code

Resulting metadata
exchange file

Generated
documentation website[1]

[1]Swagger Editor:
https://editor.swagger.io/

Next step: "FAIRification as a By-Product" Approach

A Light-Weight Approach to Develop/Revise Practices that FAIRify Artifacts as a By-Product

- Learn from SciKGTeX and DataDesc by **generalizing** their **development ideas**

- Inspired by **agile method** in software engineering
  - **Values:** Justify why we do something
  - **Principles:** Guide actions
  - **Practice:** Concrete actions to achieve values

- Core question
  - How can we ensure that actions of a practice are achieve the aspired values?
  - **Principles** help to **design** the **actions** of a practice so that the **values** are **achieved**

Practice

Concrete actions to achieve the values.

Principles

Concretize the values by providing recommendations for the actions in a practice.

Values

Abstract criteria for justifying the direction and purpose of actions in a practice.
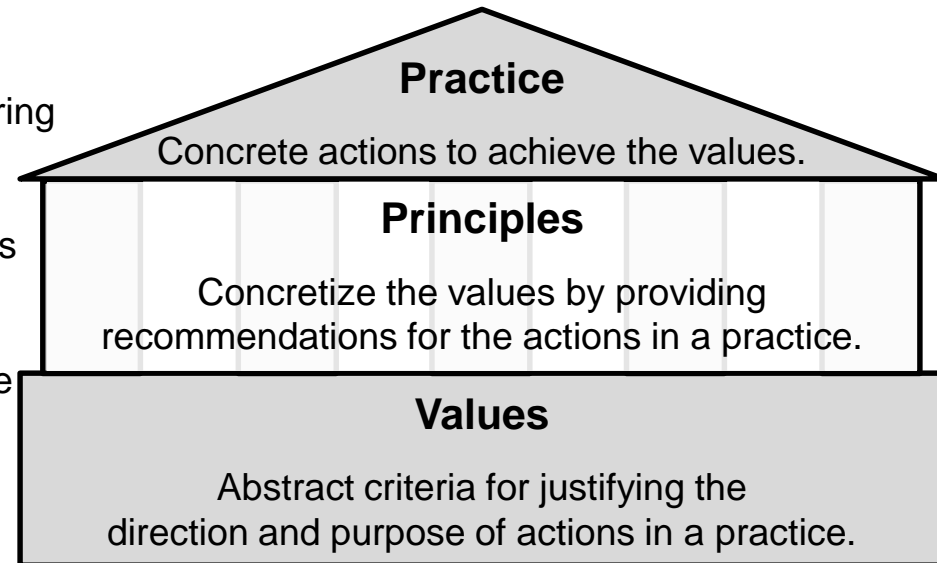
Next step: "FAIRification as a By-Product" Approach

First Draft of the Values and Principles for the "FAIRification as a By-Product" Approach

— ● Values
- **Integration:** FAIRification as integrated part
- **Involvement:** Involve only the creator
- **Simplicity:** As easy as possible in terms of process, technology, knowledge, and skills

— ● Principles
- **Focus:** One particular practice at a time
- **Creator:** The creator do the FAIRfication
- **Concurrency:** FAIRification during the practice
- **Embedding:** Embed FAIR data into the artifact
- **Fallback option:** Ensure that the artifact is created even without the FAIRification

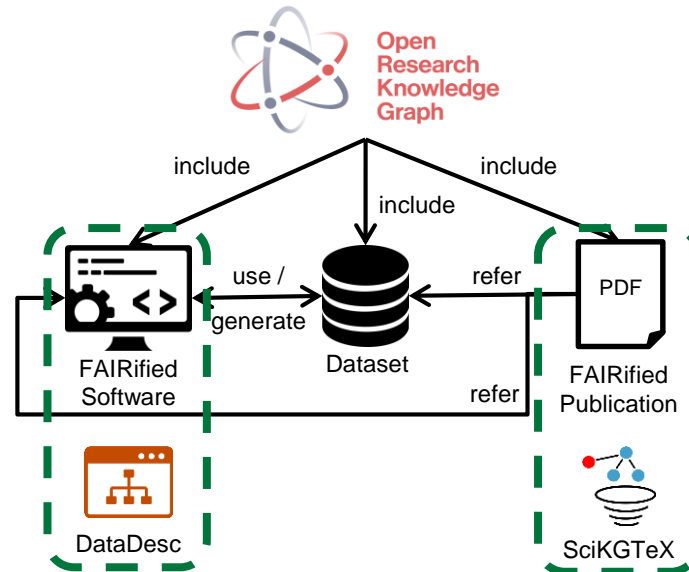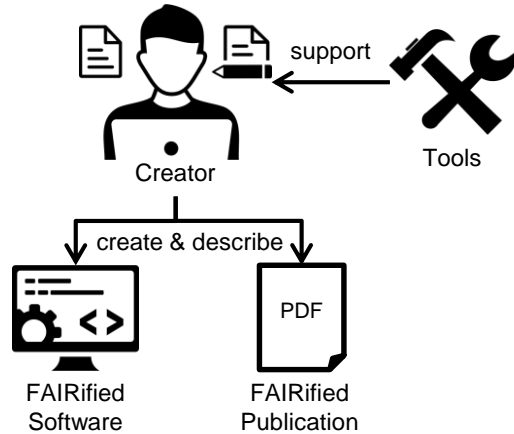| Principles	Values		
	Integration	Involvement	Simplicity
Focus	X	X	
Creator		X	X
Concurrency	X		X
Embedding	X		X
Fallback option	X		X

Conclusion

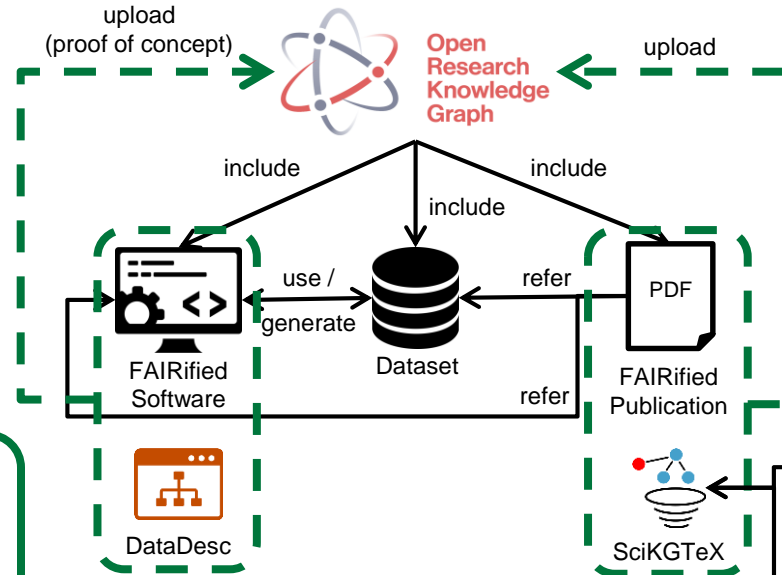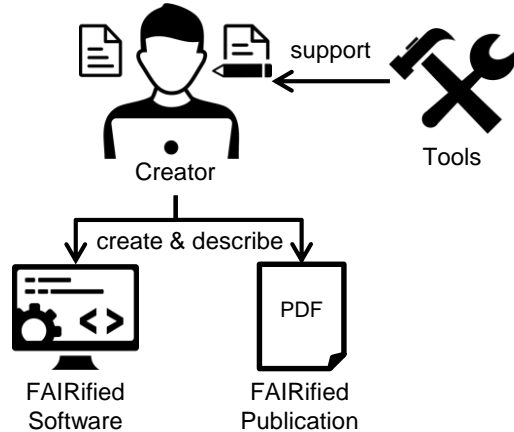**FAIRification of Artifacts at the Time of their Creation**
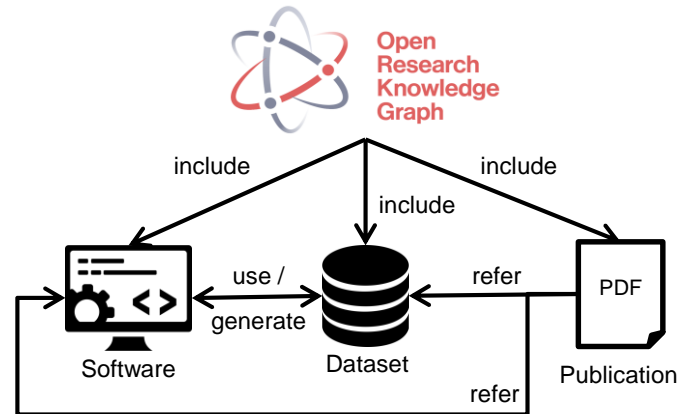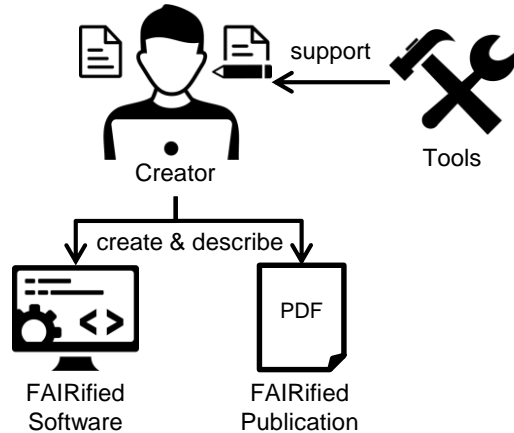
Conclusion

FAIRification of Artifacts at the Time of their Creation

Conclusion

FAIRification of Artifacts at the Time of their Creation

Ellen's Message:
Production of **"FAIR-by-Design"** **artifacts** is an opportunity to improve their **storage**, **access**, and **re-use**.