



<https://vsr.informatik.tu-chemnitz.de/research>



TECHNISCHE UNIVERSITÄT
CHEMNITZ

Helsinki, FINLAND

ICWE 2020

20TH INTERNATIONAL CONFERENCE ON WEB ENGINEERING

SolidRDP: Applying Solid Data Containers for Research Data Publishing

André Langer, Dang Vu Nguyen Hai and Martin Gaedke

VSR.Informatik.TU-Chemnitz.de

Open Data and Industrial Track
Helsinki, Finland, Online on 9 – 12 June 2020





Professorship for Distributed and Self-organizing Systems Chemnitz University of Technology, Germany



André Langer
Research Associate



Dang Vu Nguyen Hai
Research Associate

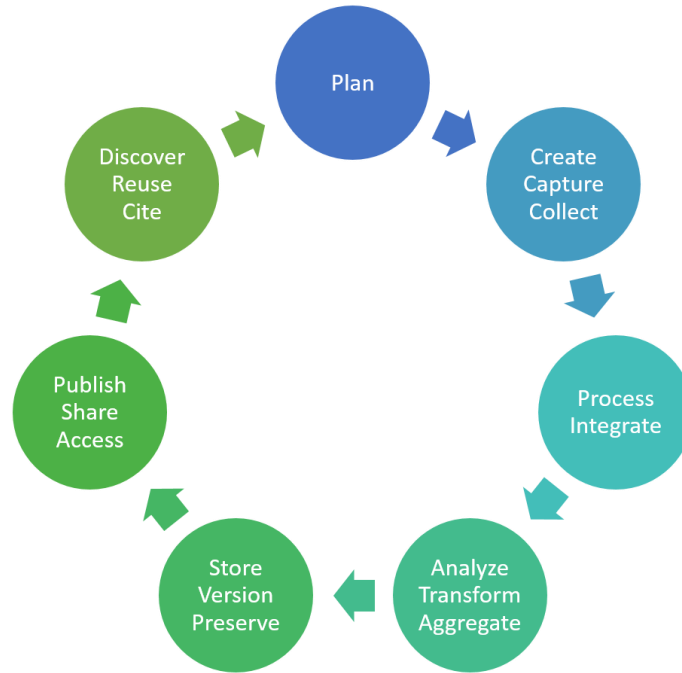


Martin Gaedke²
Full Professor



Problem Description

Research Data Management Lifecycle



PIROL: Cross-domain Research Data Publishing with Linked Data technologies

André Langer^[0000-0001-7073-5377]*

Chemnitz University of Technology, Germany
andre.langer@informatik.tu-chemnitz.de

Abstract. Effective research data management for traceability, preservation and reuse is an important part of good scientific practice and is already under discussion over a long period of time. However, the digital transformation in science also led to new challenges for researchers on how to describe, publish and share their research data. This includes the interdisciplinary annotation and discovery of research data, data privacy issues in exposure of data with trends to decentralized platforms as well as sophisticated automatisms to ensure data quality and compliance aspects. Only limited tool support exists for these processes so far. The following research project will use Linked Data principles to improve the current situation in this problem domain. It will first focus on components and services, that assist researchers in the annotation process of their research data. Next, it will investigate how this research data can be stored and discovered in decentralized, multi-user scenarios to allow data reuse under respect of data privacy concerns. In a third step, meta data descriptions will be used to apply automated data conformance and quality assessment operations on scientific data.

Keywords: Data Annotation, Data Publishing, Decentralization, Data Quality, Linked Data, SoLID, Open Research

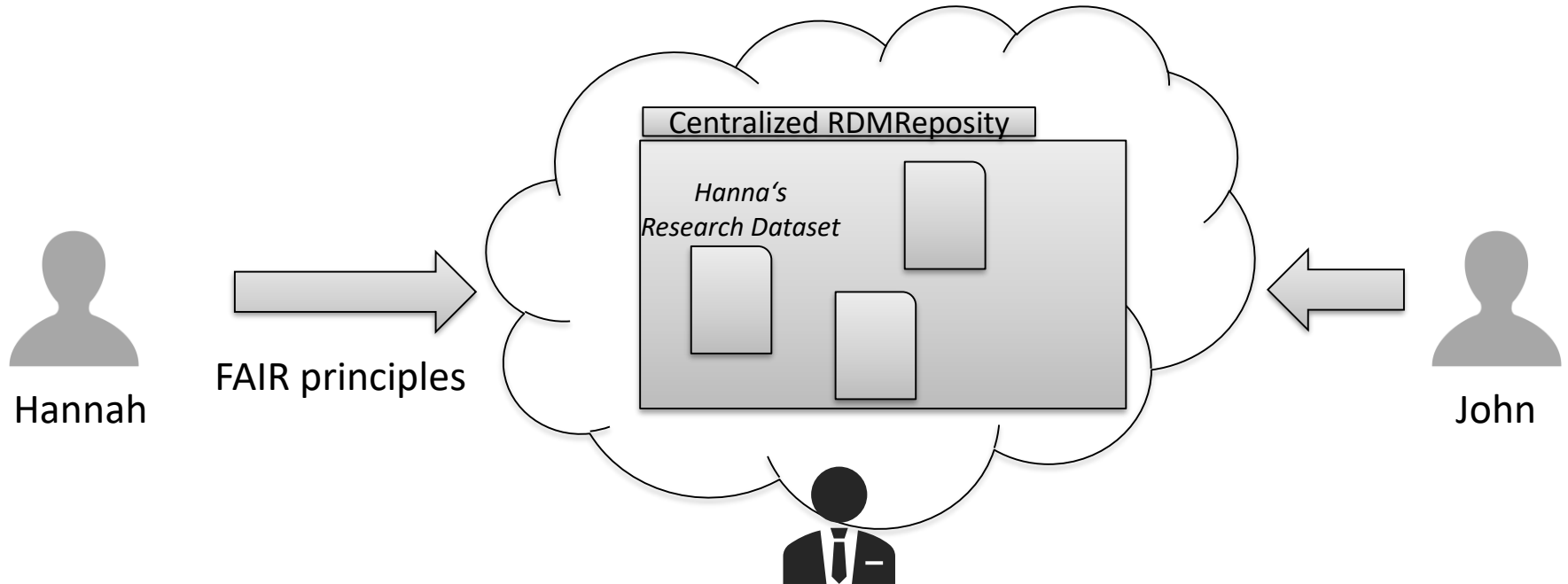
1 Introduction

The term digital research data (sometimes also referred to as scientific data or scholarly data) covers in principle any kind of digital artifact that is associated with scientific research [8].

Research data is an essential artifact of scientific work: It leads to insights, makes research reproducible and validates findings. Therefore, it is inseparable connected to the results of research and also forms the base for future activities. The digital transformation of science has risen new opportunities and issues for all involved stakeholders on national and international level on how to annotate, publish, archive, find and reuse digital research data. Researchers are confronted with new challenges when aiming to publish and share or reuse research data digitally. This applies especially for multifaceted research data where researchers from many cross-domain knowledge domains work together in a collaborative fashion.

* This research is developed under the supervision of Prof. Dr.-Ing. Martin Gaudke.

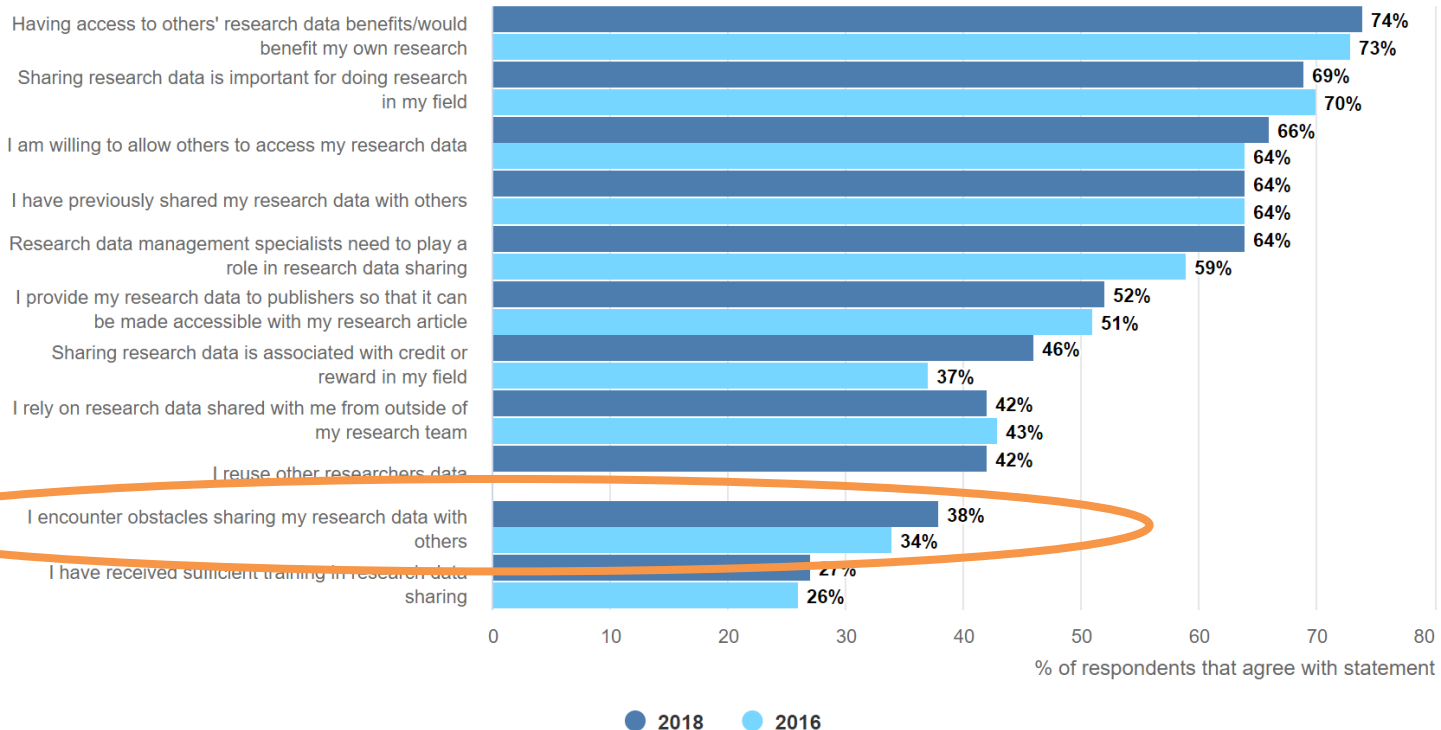
Research Dataset Publishing Nowadays



Attitudes of researchers: % of respondents that agree with statement



Reference date: 2018 and 2016

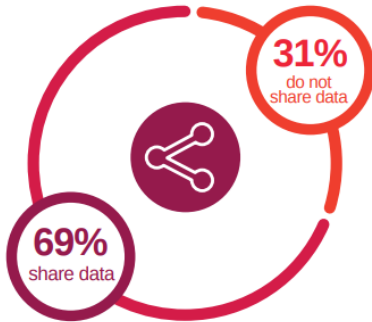


https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en

Global Data Sharing Trends*

Over 4,600 Wiley authors from 112 countries completed our 2016 Wiley Open Science Researcher Insights Survey

Data sharing in 2016



More than two thirds of Wiley researchers reported they are now sharing their data. Though this varies geographically and across research disciplines we are seeing that more researchers are sharing their data and taking efforts to make it reproducible. Archiving in institutional repositories, public repositories, and personal web pages has almost doubled since 2014.

Top 4 reasons why researchers are hesitant to share their data

- 1 **50%** - Intellectual property or confidentiality issues
- 2 **31%** - Ethical concerns
- 3 **23%** - I am concerned about misinterpretation or misuse of my research
- 4 **22%** - I am concerned that my research will be scooped

<https://authorservices.wiley.com/asset/photos/licensing-and-open-access-photos/Wiley%20Global%20Data%20Sharing%20Infographic%20June%202017.pdf>

Non-centralized Research Data Publishing

Table 1: Assessment of existing decentralized RDM solutions

	P2P-based approaches	Blockchain-based appr.	Git-based approaches
OBJ1: Ownership of data storage	o	o	+
OBJ2: Support for different formats, shapes and sizes of research datasets	+	+	+
OBJ3: Metadata integration	o	o	o
OBJ4: User support tools	+	+	+
OBJ5: Data versioning control	+	-	+
OBJ6: Data access control	-	o	o
OBJ7: Data exposure	-	o	o

+ Completely Fulfilled, o Partially Fulfilled, - Not Fulfilled



Re-decentralizing the Web, for good this time

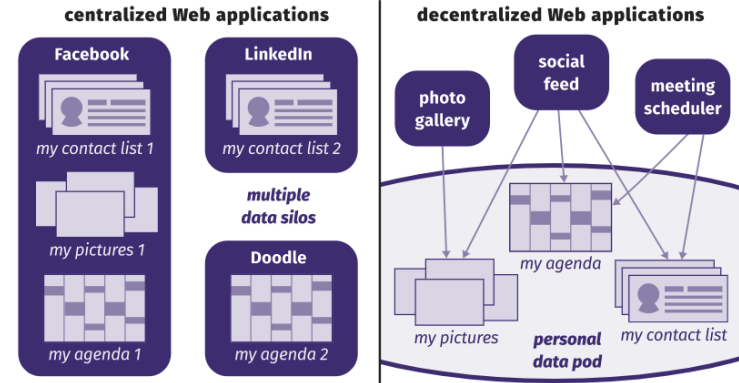
Ruben Verborgh, Ghent University – imec – IDLab

11 January 2019

ORIGINALLY DESIGNED AS A DECENTRALIZED ECOSYSTEM, THE WEB HAS UNDERGONE a significant centralization in recent years. In order to regain control over our digital self, over the digital aspects of our lives, we need to understand how we arrived at this point and how we can get back on track. This chapter explains the history of decentralization in a Web context, and details Tim Berners-Lee's role in the continued battle for a free and open Web. The challenges and solutions are not purely technical in nature, but rather fit into a larger socio-economic puzzle, to which all of us are invited to contribute. Let us take back the Web for good, and leverage its full potential as envisioned by its creator.

<https://ruben.verborgh.org/articles/redecentralizing-the-web/>

You can grant apps and people access to very specific parts of your data.



<https://rubenverborgh.github.io/Solid-DeSemWeb-2018/>



SOLID

Use Solid For Developers For Enterprises FAQ

Switch between storage and apps while taking the data along

<https://solidproject.org/>





How can we adapt a **Linked Data Platform** based approach to publish Research Datasets?

Objectives

OBJ1: Ownership of data storage

OBJ2: Support for different formats, shapes and sizes

OBJ3: Metadata Integration

OBJ4: Ease of User Interface and Support

OBJ5: Data Versioning Control

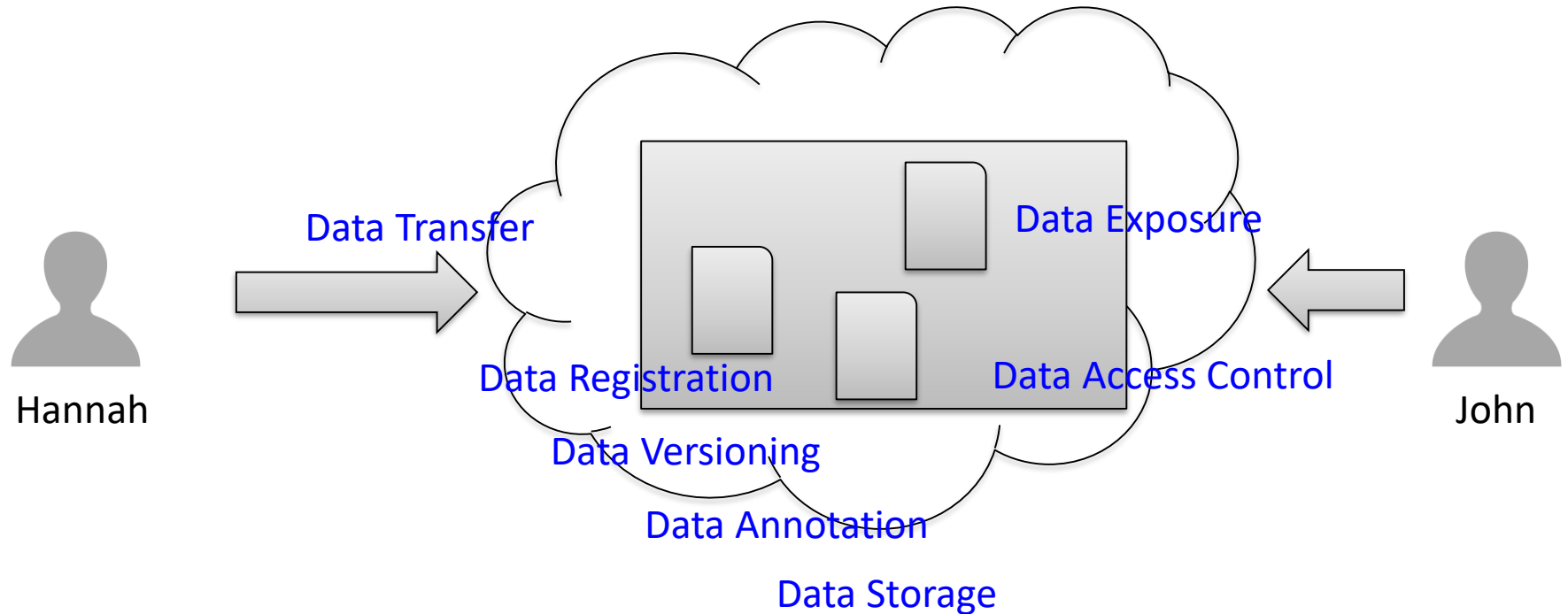
OBJ6: Data Access Control

OBJ7: Data Exposure

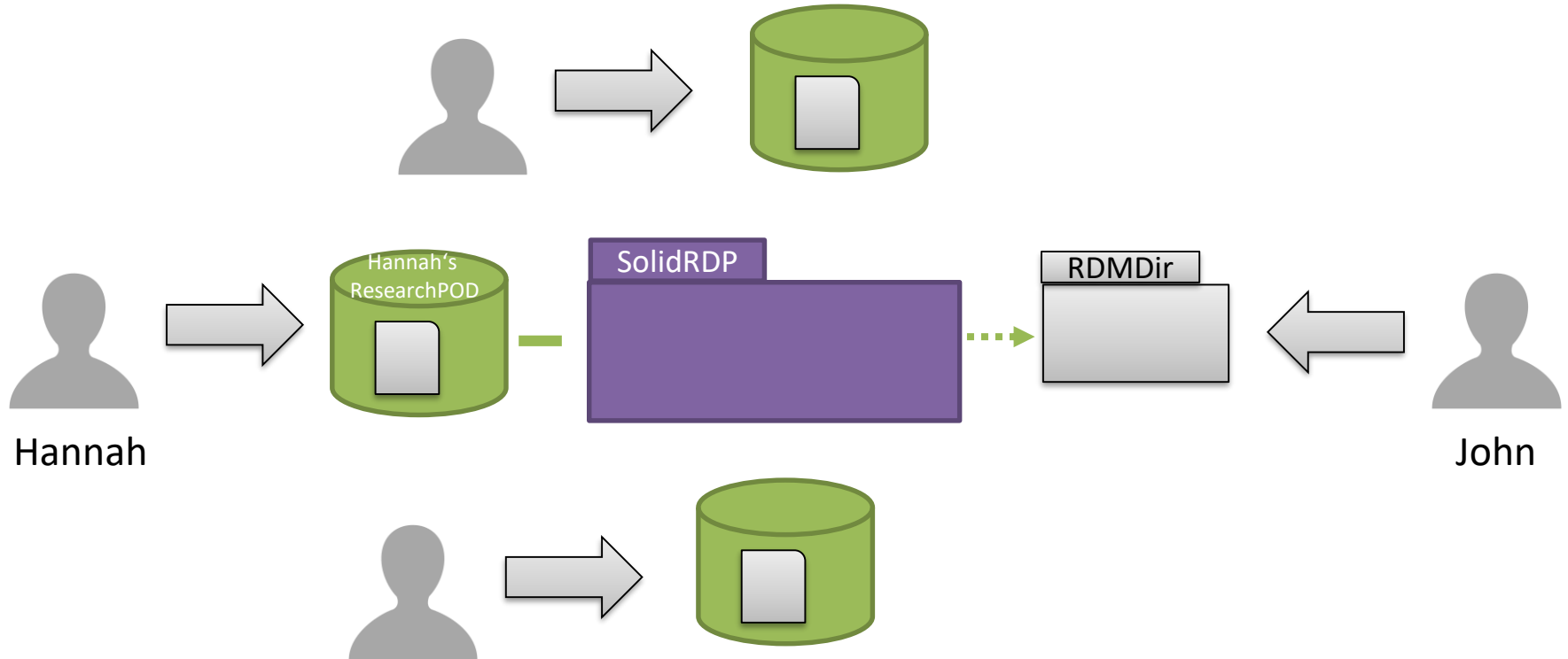


Approach

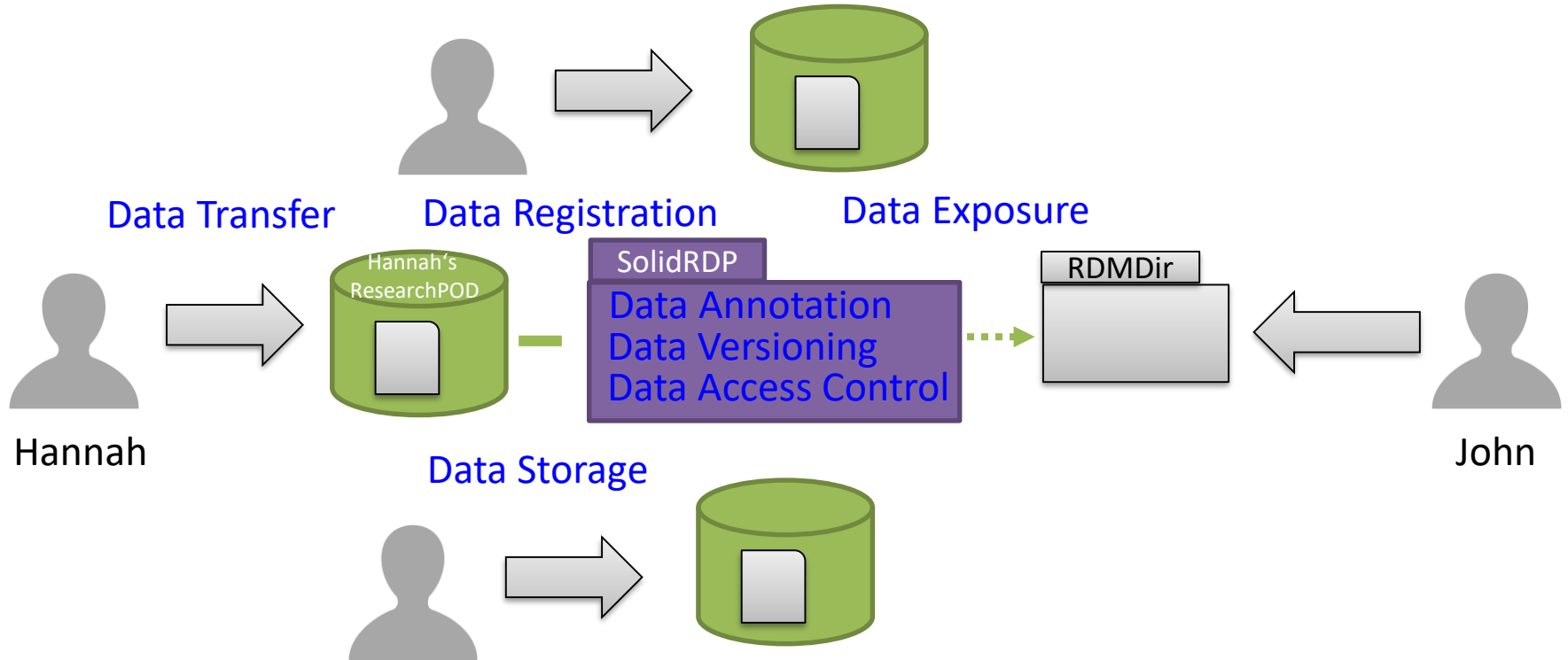
Research Dataset Publishing Nowadays



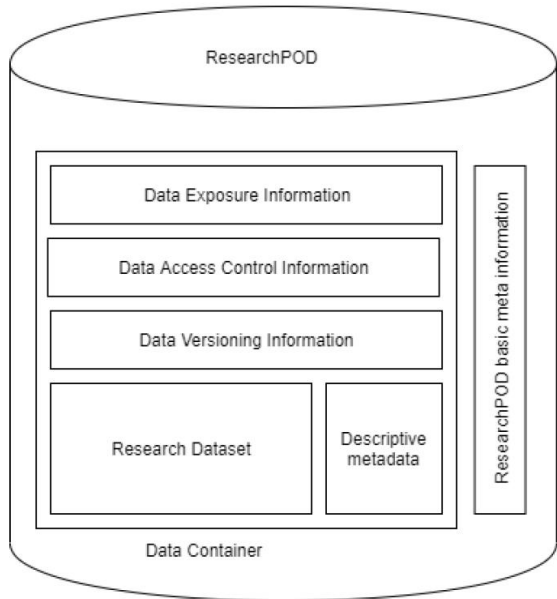
Research Dataset Publishing Decentralized



Research Dataset Publishing Decentralized

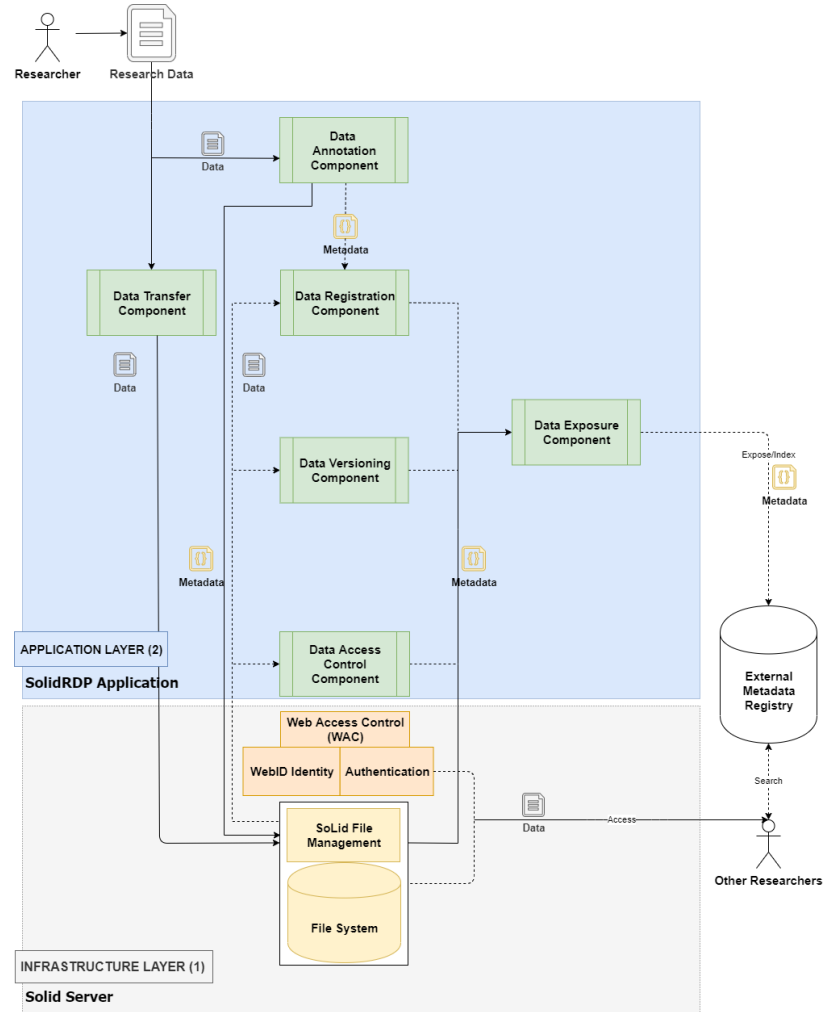


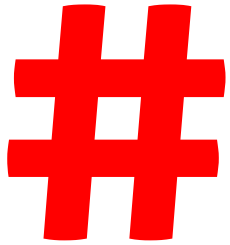
SolidRDP Approach



Structure of a DataContainer within a ResearchPOD

SolidRDP Architecture





SolidRDP PoC



<https://purl.net/net/vsr/solidrdp>

Evaluation Strategy

1. **PoC** with ***SolidRDP*** as prototypical implementation
2. **Lab Study** with Systematic Observation and Questionnaire
3. **Objectives – based Evaluation**



Conclusion

Contribution

- **Conceptual** Model for Research Data Publishing
- **Data Container** – based approach in ResearchPOD
- **Component-based** *SolidRDP* web application

Future Work

- **Investigation** of each involved Component in detail
- **Discovery** of Research Data in ResearchPODs
- **Integration** in existing research environments



<https://vsr.informatik.tu-chemnitz.de/research>

Inspired and Interested?

Andre.Langer@Informatik.TU-Chemnitz.de

@myVSR

 /myVSR



<https://vsr.informatik.tu-chemnitz.de/research>