Slides: https://osf.io/4zyfg/
OSF Project: https://osf.io/7n5yp/

FIP for Purpose

ENVRI-FAIR's FAIR assessment number 3

- Session 1, January 25 (9:00-12:00 CET): FIP review & revision
- Session 2, January 28 (9:00-12:00 CET): Consultation
- **Session 3, February 22 (9:00-12:00 CET): Convergence**

Erik Schultes
Barbara Magagna

GO FAIR foundation

CODEVENCE

umweltbundesamt
ENVIRONMENT AGENCY AUSTRIA

VU VRIJE UNIVERSITEIT AMSTERDAM

ENVRI-FAIR, Jan. 25&28, Feb. 22 '22 ~ www.osf.io/7n5yp/

CC BY NC SA

FIP Wizard & Convergence Team

October 2021 - February 2022

Barbara Magagna

Marek Suchánek

Jacintha Schultes

Erik Schultes

Tobias Kuhn

ENVRI-FAIR, Jan. 25&28, Feb. 22 '22 ~ www.osf.io/7n5yp/

CC BY NC SA

Common Notes:

https://docs.google.com/document/d/1XBnGDL9YZ9yZle25oUupJD-0d2AXwkKo6lw3Jvsu3aA/edit?usp=sharing

FIP Results:

https://docs.google.com/spreadsheets/d/1jg9uV2UiK7DpRVvXIRklN6X-H_TNPvu9oy91E8E1cBI/edit?usp=sharing

3nd event: Convergence Session February 22 - 3 hours

- 9:00 Welcome
- 9:05-10 min BM: Brief history of ENVRI-FIPs (3rd assessment)
- 9:15-20 min EAS: What is Convergence?
- 9:35-15 min BM: Data analysis on FIPs
- 9:50-10 min Break
- 10:00-60 min EAS: Review with a focus on convergence strategies
- 11:00-20 min BM: Presentation of FIPs / convergence matrix / discussion
- 11:20-20min EAS: Outlook - CODATA / FDO Forum

Brief history of ENVRI-FIPs (3rd assessment)

FIP for Purpose

Session 1, January 25: FIP review & revision

Session 2, January 28: Consultation

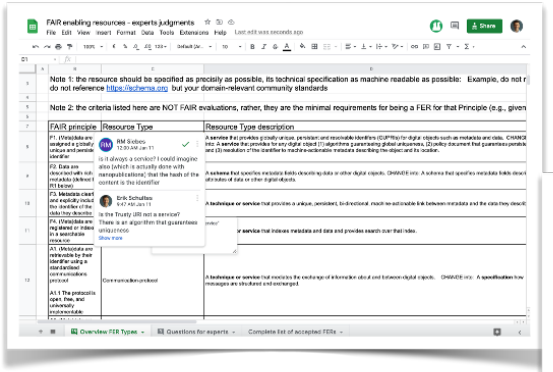
FIP for Convergence

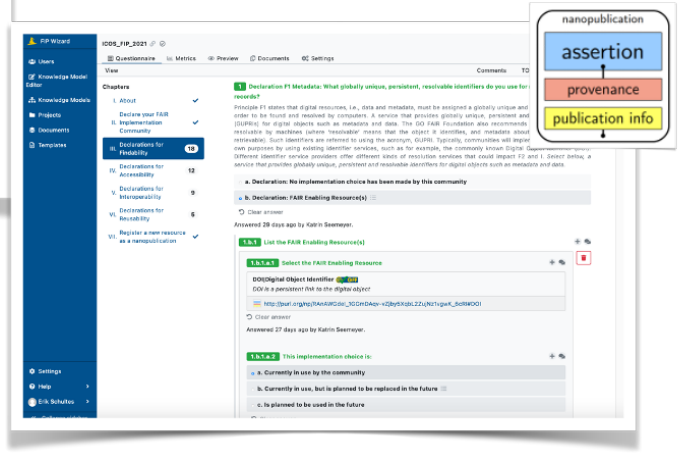
Session 3, February 22: Convergence

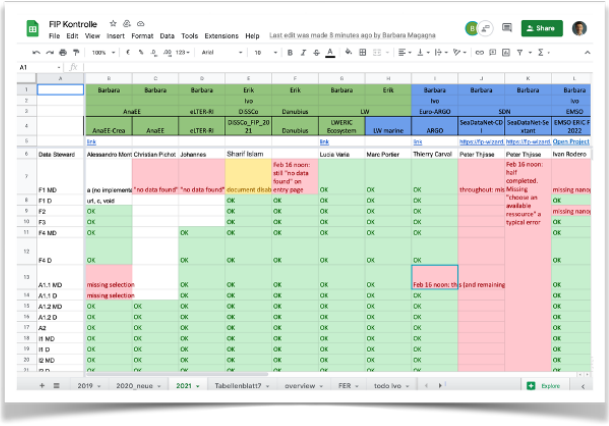
9:00 – 10:00	Introduction
	Convergence
	Process
10:00 – 11:00	Convergence Details and Strategies
11:00 – 12:00	Review presentation and Outlook

ENVRI-FAIR, Jan. 25&28, Feb. 22 '22 ~ www.osf.io/7n5yp/

FIP questionnaire and output

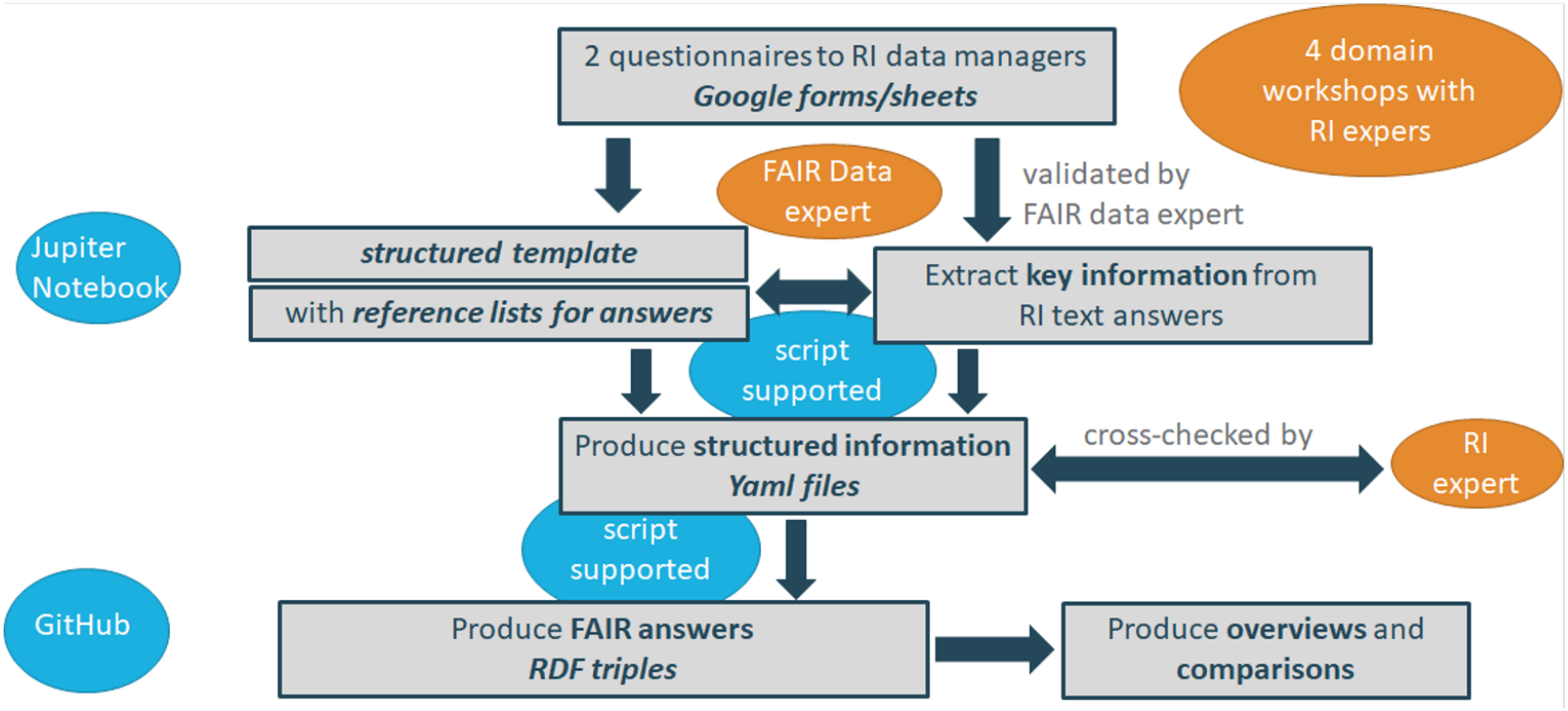






ENVRI-FAIR, Jan. 25828, Feb. 22 '22 ~ www.osf.io/7n5yp/

2019: Questionnaire based assessment



ENVRI-FAIR, Jan. 25828, Feb. 22 '22 ~ www.osf.io/7n5yp/

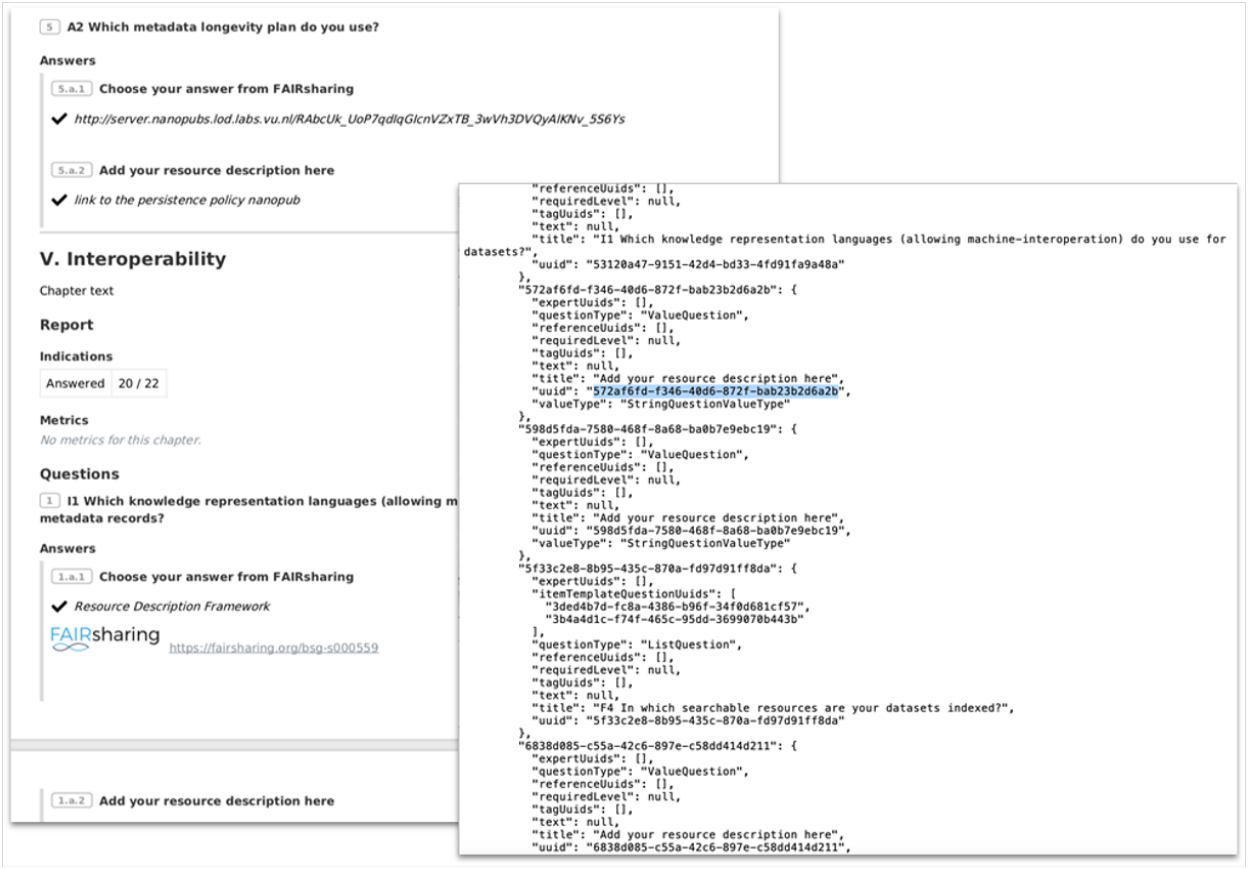
2019: from human (text) to machine readable information (yaml)

1.10 Do you assign PIDs manually or automatically?	The Argo DOI fragments are assigned automatically	automatically	
1.11 Which PID registration provider do you use?	DataCite	SEANOE	
1.12 Do you use the PID Record to store attributes about the data?	"Yes for the monthly snapshot (the DOI+ fragment) No otherwise. "	yes	
1.13 Are these repositories certified? If so, which methods are used?	"Yes, Ifremer is DSA and IODE certified. Ifremer-Sismer is in certification process as "RDA-Trustworthy repository" "	- Data Seal of Approval	
1.14 Are repository policies mentioned at the website? If so, indicate the major ones.	"Yes https://creativecommons.org/licenses/by/4.0/ "	- data access	
1.15 Are your repositories registered in a registry? If so which registry?	Yes, GEO registry	GEO	
1.16 Which persistency guaranties are typically given?	"The Argo long term archive is managed by US-NCEI. US-NCEI has a Unesco-WMO mandate as world data centre (WDC-A). "	NULL	
1.17 Which are the most popular data types used?	The self-describing NetCDF CF format Argo implementation	binary	

```yaml
repositories:
- URL: http://doi.org/10.17882/42182
  name: Euro-Argo Data
  kind: data repository
  data repository type: central
  metadata repository type: central
  software: NetCDF file
  identifier:
  - IRI: http://doi.org/10.17882/42182
    persistency-policy document IRI: https://doi.org/10.13155/44515
    kind: DOI
    system: SEANOE
    assigned: automatically
    provider: SEANOE
    includes-attributes: yes
  certification methods:
  - Data Seal of Approval
  - IODE certified
  - RDA Trustworthy Data Repository
  policies:
  - data access
  registries:
  - GEO
  persistency-guaranty: NULL
  access mechanisms:
    authentication method: VOID
    access protocol URL: https://doi.org/10.17882/42182
```

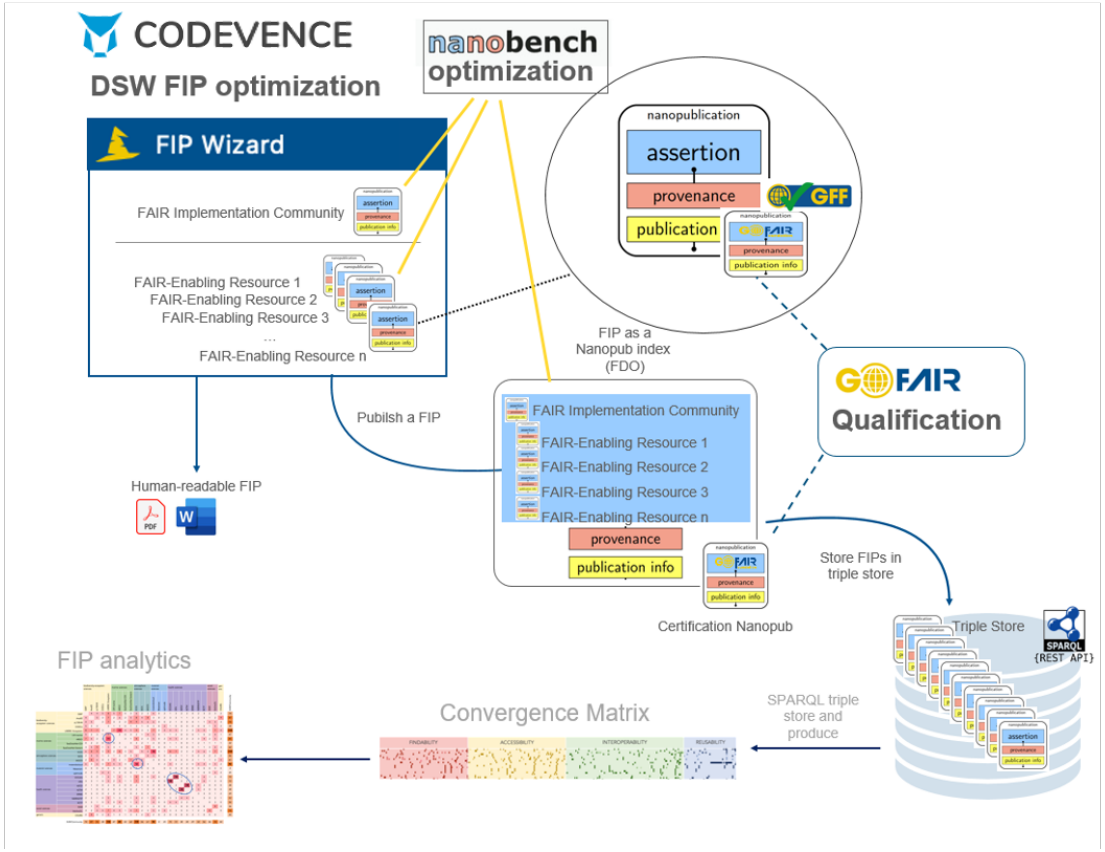
ENVRI-FAIR, Jan. 25828, Feb. 22 '22 ~ www.osf.io/7n5yp/

FIP Outputs



ENVRI-FAIR, Jan. 25828, Feb. 22 '22 ~ www.osf.io/7n5yp/

Human – machine - human



ENVRI-FAIR, Jan. 25828, Feb. 22 '22 ~ www.osf.io/7n5yp/

What is Convergence?

https://doi.org/10.3390/info12110472

information

MDPI

Article

Revolutions Take Time

Peter Wittenburg [1,*] and George Strawn [2]

[1] FDO Forum, 2333 CR Leiden, The Netherlands
[2] National Academy of Sciences, Washington, DC 20001, USA; gostrawn@gmail.com
* Correspondence: peter.wittenburg@mpcdf.mpg.de; Tel.: +49-282149180

Abstract: The 2018 paper titled "Common Patterns in Revolutionary Infrastructures and Data" has been cited frequently, since we compared the current discussions about research data management with the developments of large infrastructures in the past believing, similar to philosophers such as Luciano Floridi, that the creation of an interoperable data domain will also be a revolutionary step. We identified the FAIR principles and the FAIR Digital Objects as nuclei for achieving the necessary convergence without which such new infrastructures will not take up. In this follow-up paper, we are elaborating on some factors that indicate that it will still take much time until breakthroughs will be achieved which is mainly devoted to sociological and political reasons. Therefore, it is important to describe visions such as FDO as self-standing entities, the easy plug-in concept, and the built-in security more explicitly to give a long-range perspective and convince policymakers and decision-makers. We also looked at major funding programs which all follow different approaches and do not define a converging core yet. This can be seen as an indication that these funding programs have huge potentials and increase awareness about data management aspects, but that we are far from converging agreements which we finally will need to create a globally integrated data space in the future. Finally, we discuss the roles of some major stakeholders who are all relevant in the process of agreement finding. Most of them are bound by short-term project cycles and funding constraints, not giving them sufficient space to work on long-term convergence concepts and take risks. The great opportunity to get funds for projects improving approaches and technology with the inherent danger of promising too much and the need for continuous reporting and producing visible results after comparably short periods is like a vicious cycle without a possibility to break out. We can recall that coming to the Internet with TCP/IP as a convergence standard was dependent on years of DARPA funding. Building large revolutionary infrastructures seems to be dependent on decision-makers that dare to think strategically and test out promising concepts at a larger scale.

Keywords: data management; data infrastructures; FAIR principles; FAIR Digital Objects

check for updates

Citation: Wittenburg, P.; Strawn, G. Revolutions Take Time. *Information* **2021**, *12*, 472. https://doi.org/10.3390/info12110472

Academic Editor: Riccardo Albertoni

Received: 10 October 2021
Accepted: 10 November 2021
Published: 16 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

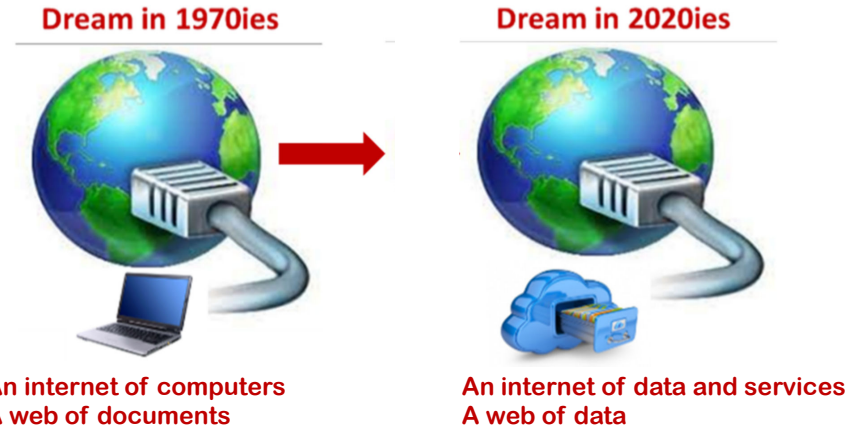
Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and

1. Introduction

In our 2018 paper "Common Patterns in Revolutionary Infrastructures and Data" [1], we argued that revolutionary infrastructures can be characterized by simple core standards which, on the one hand, promise some stability associated with a step of disruption but, on the other hand, act as a new common platform for dynamic evolution above that platform. We also argued that the emerging distributed data infrastructure will be such a revolutionary infrastructure and thus requires a convergence towards such a simple standard as its key pillar. We identified the FAIR principles [2] and the FAIR Digital Objects (In this paper we will not describe the FDO concept in detail but refer to some publications and the Base Definition which is published at the fairdo.org website (http://fairdo.org; accessed on 15 November 2021). In short one can state that an FDO has a structured

Table 1. This table indicates a rough categorization of phases of virtual integrations in IT development.

1950s	many individual computers	separated data sets
1990s	one virtual computer	separated data sets
2030s	one virtual computer	one virtual data collection



https://doi.org/10.23728/B2SHARE.4E8AC36C0DD343DA81FD9E83E72805A0

Common Patterns in Revolutionary Infrastructures and Data

Peter Wittenburg, Max Planck Computing and Data Facility
George Strawn, US National Academy of Sciences
February 2018

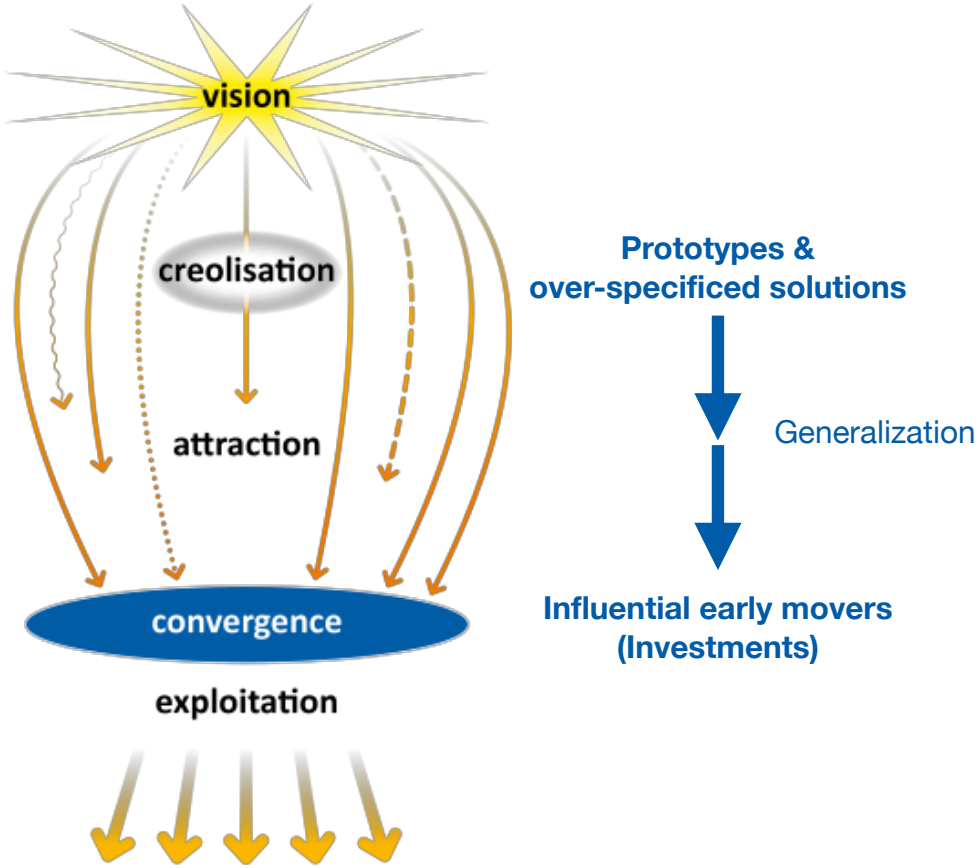
1. Summary

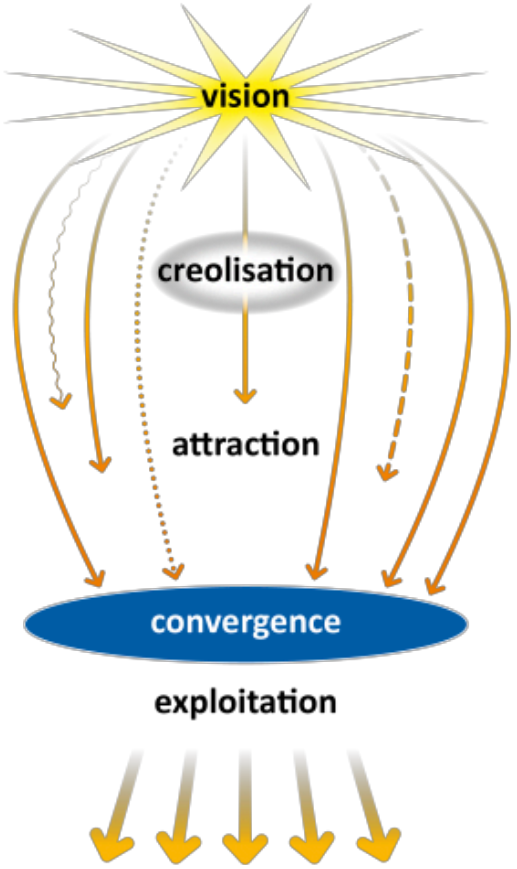
Societies have seen large infrastructures emerge when new technologies become available. From history we see that such infrastructures can have a huge influence on all aspects of societal life. Moreover, some patterns appear to reoccur in the evolution of such infrastructures. *Early visions* about the possibilities of a new technology lead to a phase of *creolization*[1] of approaches resulting in a deeper knowledge of the technology's pros, cons and limitations. A huge "solutions space" emerges and fragmentation results. Some solutions are more *attractive* than others, but a final phase transition occurs where the experts converge towards broadly accepted principles and specifications that lead to *exploitation and standardization*.

It appears that the "data infrastructure" is evolving into such a large infrastructure, with a potentially large influence on societies, industry and science. In order to gain new insights about complex relationships in nature, societies and minds, by integrating data from different silos we have seen an explosion of (non-interoperable!) solutions for data management, access and processing, i.e. we have entered a phase of *creolization*. Also, we have an increasingly clear view of the current inefficiencies in working with data. These inefficiencies retard innovation and broad participation, which will become even more important as billions of smart devices produce the data deluge of the Internet of Things. Stakeholders have begun looking for steps toward *convergence* that would increase efficiency without hampering innovation.

Comparing the evolution of the data infrastructure with the evolution of the infrastructures of electrification, computer networking and of information networking (WWW), we can observe that, despite all initiatives already taken, we have not reached convergence on a set of universals that would boost developments and create a momentum towards an efficient and interoperable data infrastructure. We propose that such a set of universals could be based on the concepts of "Digital Objects" (DOs), persistent identifiers (PIDs), and metadata (including data typing). These concepts could greatly reduce current inefficiencies in data processing and open the way towards automatic processing. In particular, the Core Data Model of the Research Data Alliance (RDA) provides a design for a universal Digital Object Access Protocol (DOAP, comparable to IP for the Internet or HTTP for the Web) which can interconnect the many organizations of data in use today, such as cloud systems, files systems, SQL databases, no-SQL databases and so forth. The agreement on fairly simple but potentially universal commonalities such as PIDs, DOs, and a DAOP could create the confidence for many developers to invest in data infrastructure building. We believe that it is time to take this step towards convergence.

Acknowledgements
We would like to acknowledge the many contributions to this discussion from close collaborators during the last months. In particular, we should mention here Robert Kahn, Larry Lannom, Tobias Weigel, Barend Mons and various colleagues from the Research Data Alliance (RDA)[2] and the C2CAMP[3] initiatives.



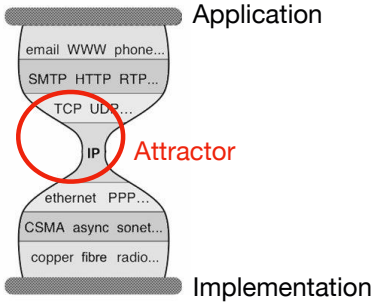


Interoperation of computer networks - an “internet”

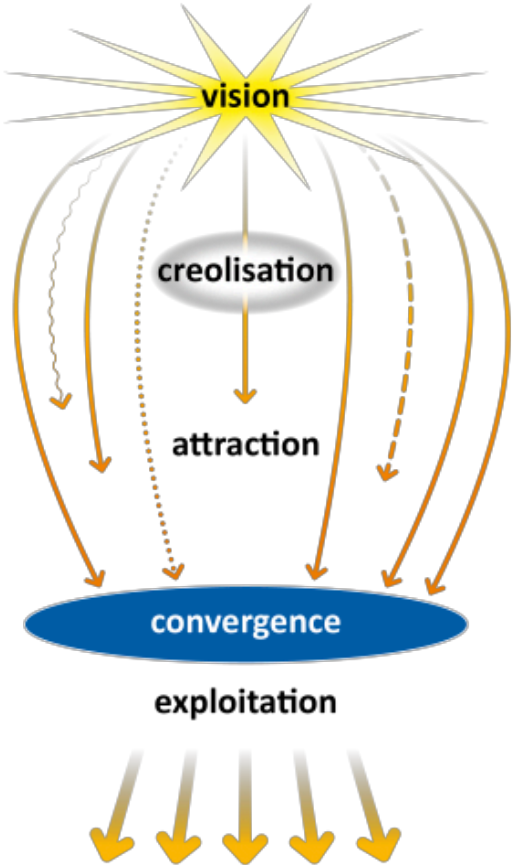
TCP/IP

1995 - private sector provided the critical mass

A recipe for convergence



- Minimal standard
- Voluntary participation
- Critical mass of users
- Rough consensus, running code
- 2 implementations of a spec

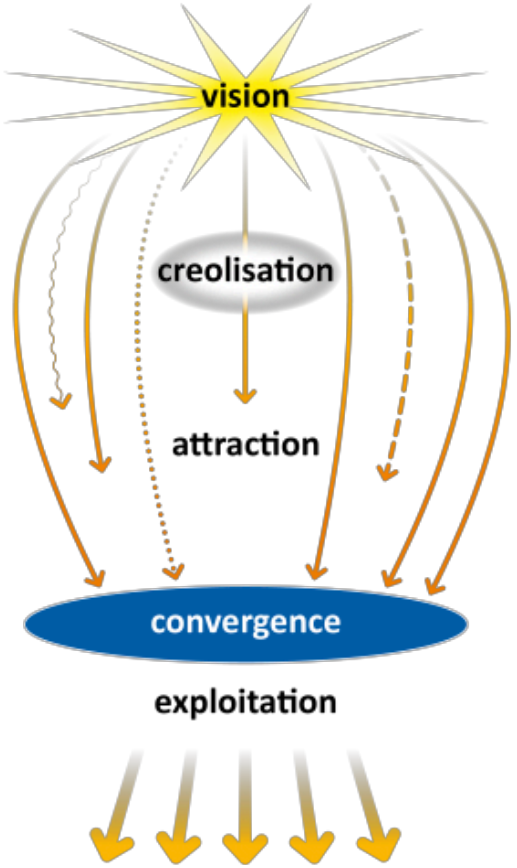


Interoperation of computer networks - an “internet”

TCP/IP

1995 - private sector provided the critical mass

Dream in 2020ies

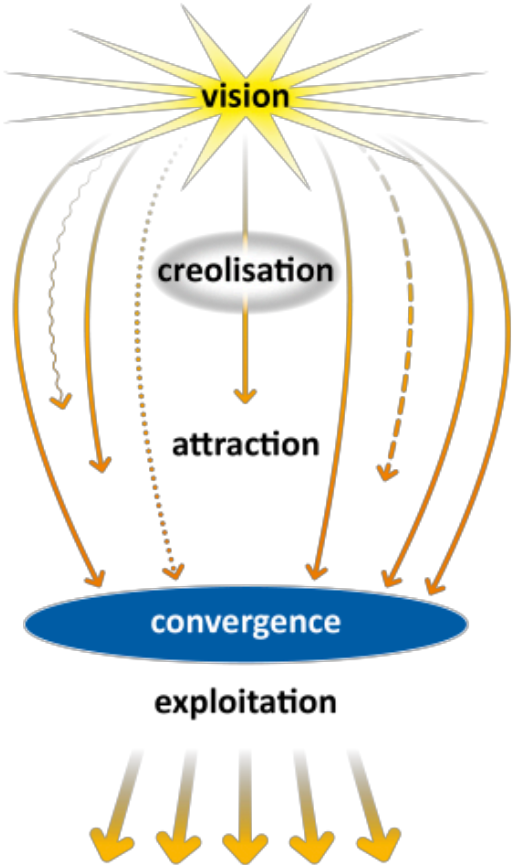


Prototypes &
over specificed solutions

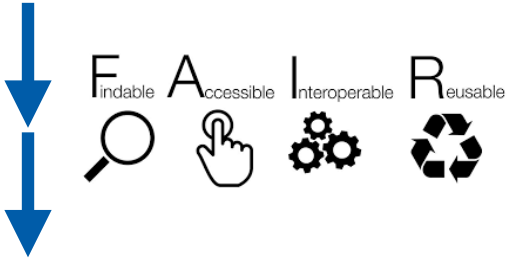
Generalization

Influential early movers
(Investments)

Dream in 2020ies

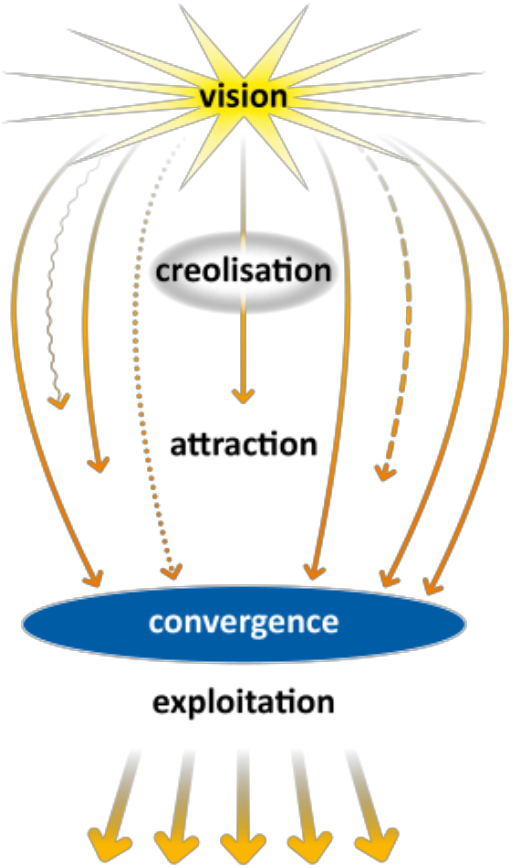


Prototypes & over specificed solutions

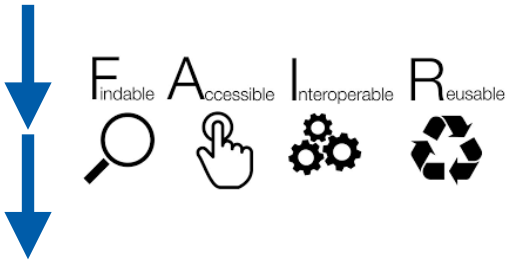


Influential early movers (Investments)

Dream in 2020ies

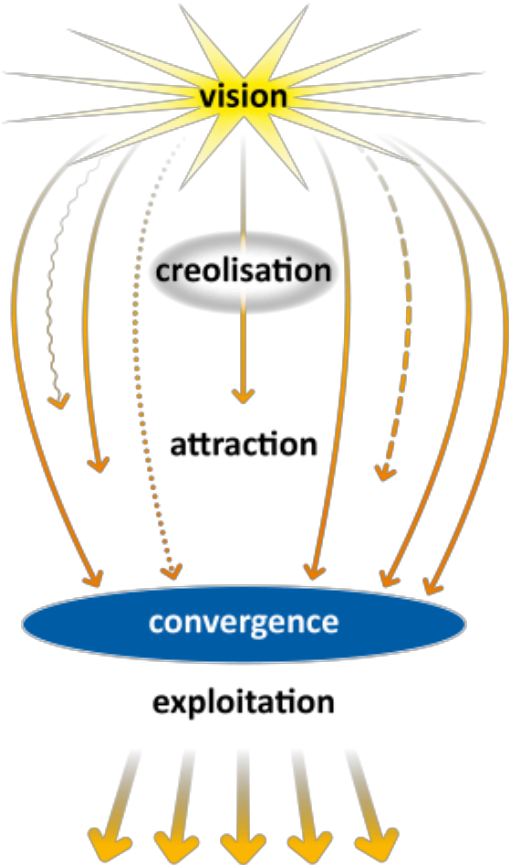


Prototypes &
over specificed solutions

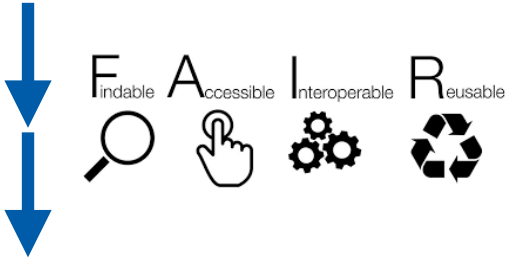


Influential early movers
(Investments)

Dream in 2020ies



Prototypes &
over specificed solutions



Influential early movers
(Investments)

ENVRI
FAIR

IDSA

FAIR DIGITAL OBJECTS FORUM

Dream in 2020ies

FIPs as a catalyst

vision

creolisation

attraction

convergence

exploitation

Prototypes & over specificed solutions

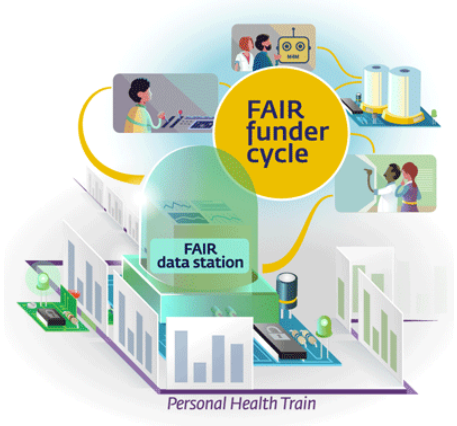
Findable Accessible Interoperable Reusable

Influential early movers (Investments)

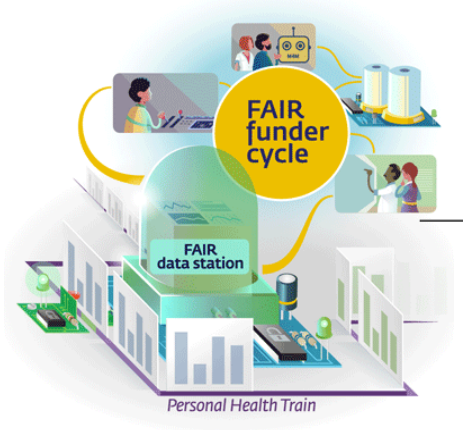
ENVRI FAIR

FAIR DIGITAL OBJECTS FORUM

Convergence on FAIR “data visting” by FIP repurposing



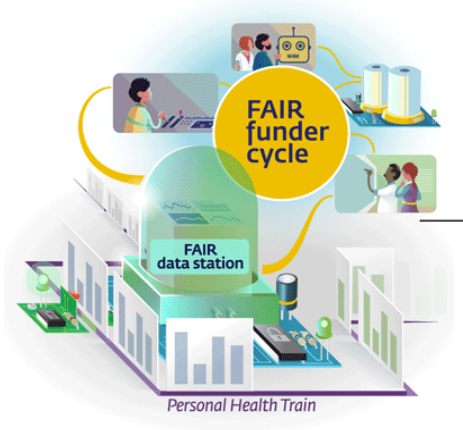
Convergence on FAIR “data visting” by FIP repurposing



FAIR Data Point
Working Draft, 23 August 2021
CEDAR BioPortal

VODAN AFRICA

Convergence on FAIR “data visting” by FIP repurposing



FAIR Data Point
Working Draft, 23 August 2021
CEDAR BioPortal

VODAN AFRICA

+ health RI
PORTAL

ZonMw COVID-19 Program

Convergence on FAIR “data visting” by FIP repurposing

FAIR Data Point
Working Draft, 23 August 2021
CEDAR BioPortal

VODAN AFRICA

+ health RI PORTAL

ZonMw COVID-19 Program

+ HL7 FHIR

TWC Trusted World Of Corona

Convergence on FAIR “data visting” by FIP repurposing

FAIR Data Point
Working Draft, 23 August 2021
CEDAR BioPortal

VODAN AFRICA

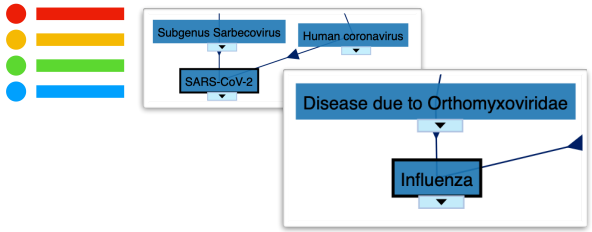
+ health RI PORTAL

ZonMw COVID-19 Program

+ HL7 FHIR

TWC Trusted World Of Corona

INCENTIVE



From FIP to Practice

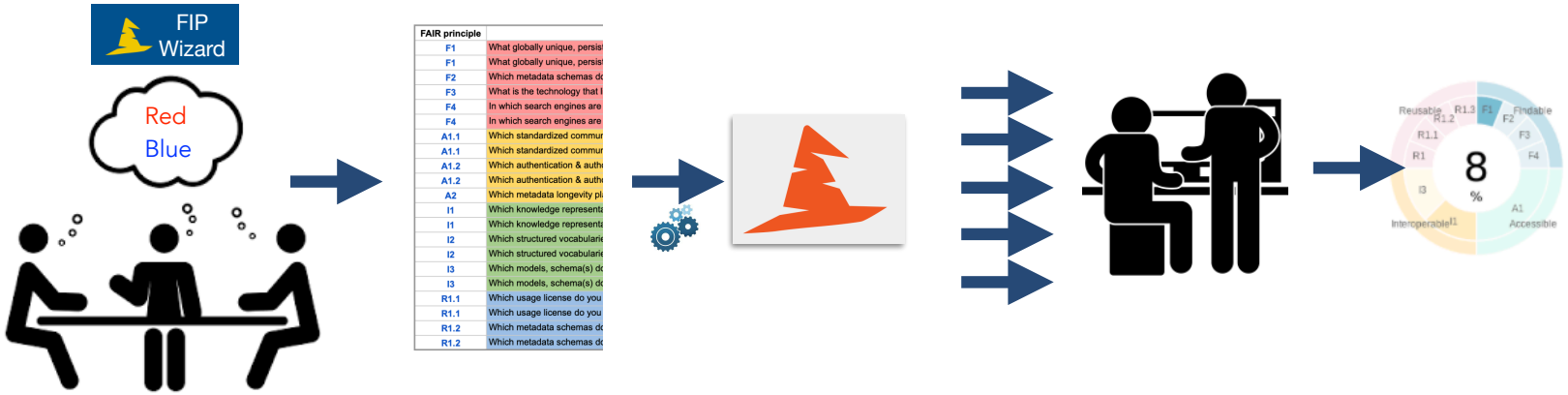
FIP workshops sponsored by funding agency

Qualified / Certified FAIR Implementation Profile

Auto-informed FAIR data management plan (template)

FAIR data management implemented in a project

FAIR Evaluation



DSW
DATA STEWARDSHIP WIZARD

DSW @ maDMP hackathon (Grand Finale)

Mapping the FIP onto the DMP
Kristina Hettne, Leiden University Libraries, June 2020

https://github.com/RDA-DMP-Common/hackathon-2020/blob/master/results.md
https://docs.google.com/presentation/d/1h7iTS9gW8A-bOkSjR3qFNPzaU0NQfGJFRtCHZuC78FY/edit?usp=sharing