Profiling ENVRIs' FAIR Implementations via FIPs to drive convergence

Barbara Magagna

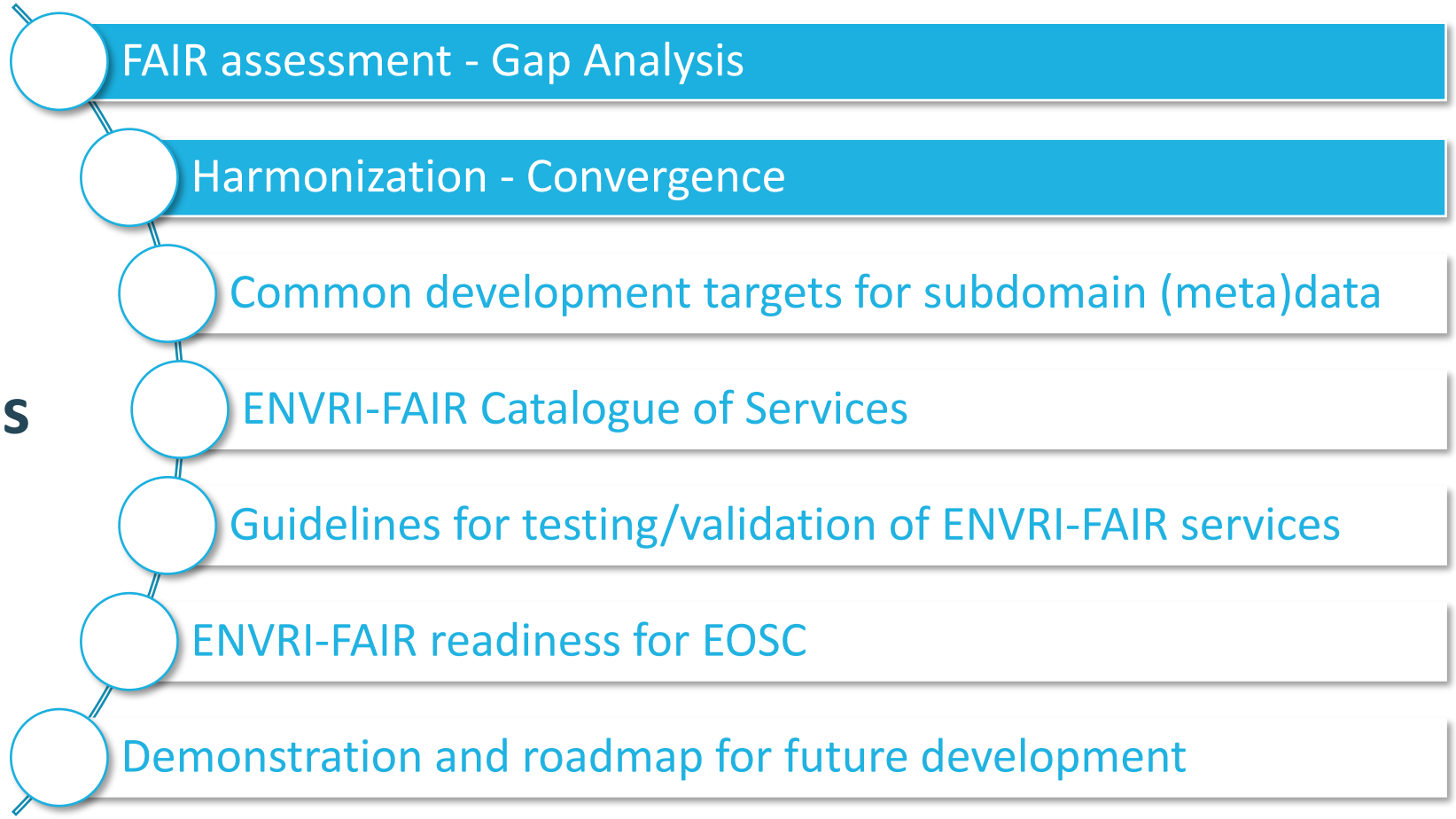
FAIR Convergence Symposium, Leiden 26 October 2022

ENVRI-FAIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

FAIR Assessment

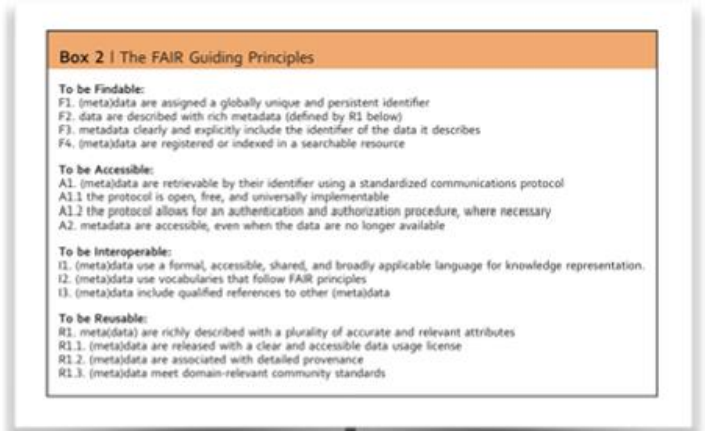
- 14 Research Infrastructures - 4 Subdomains
- Different stages of FAIRness maturity

WP5 Goals



The FAIR Implementation Profile (FIP) concept

FAIR Principles



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

questions related to the FAIR Principles

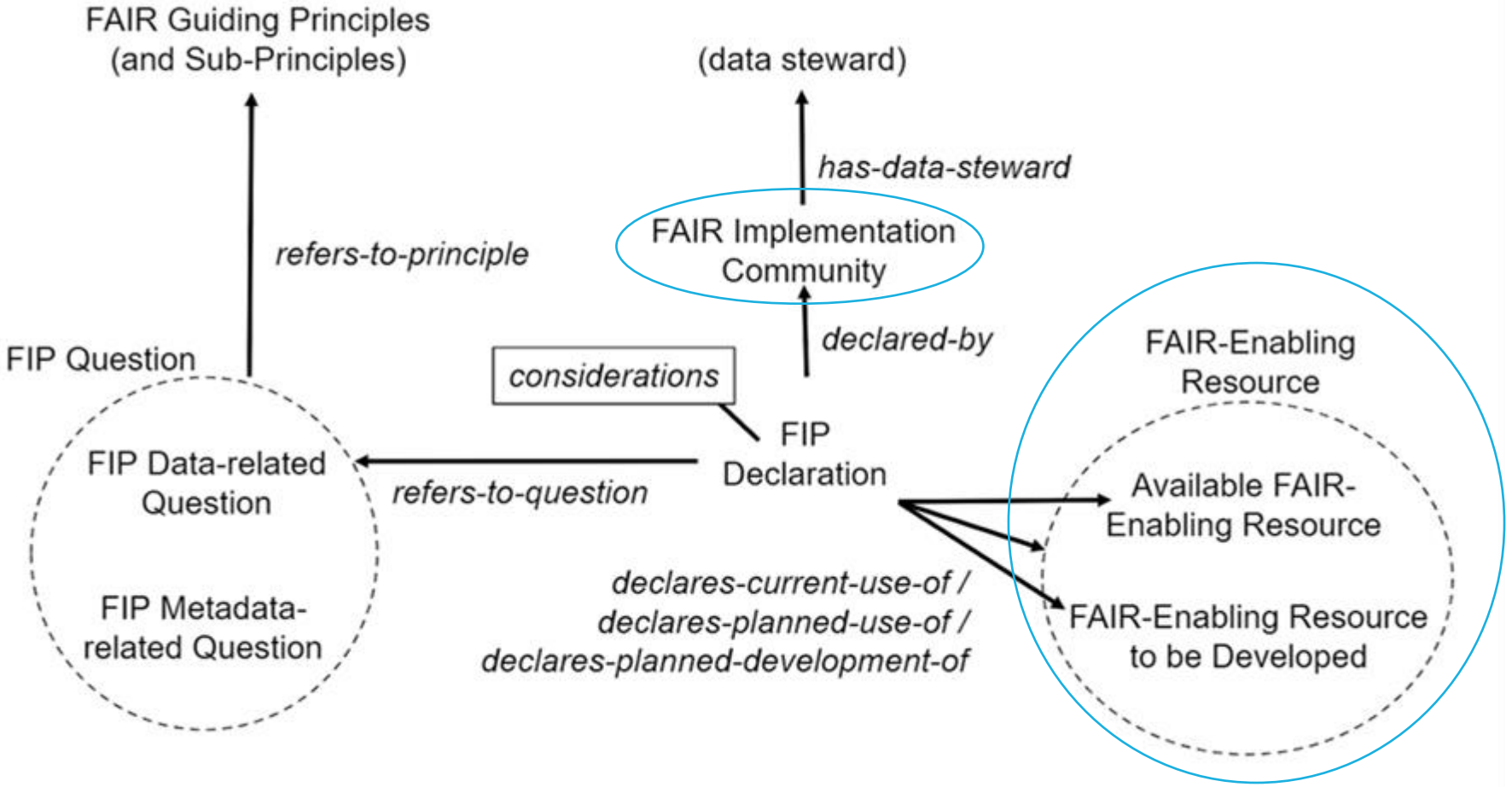
FAIR enabling resource (FER)

digital objects that provide functions needed to achieve some aspect of FAIRness

FAIR Implementation Profile (FIP)

a collection of machine-readable human agreements addressing each of the FAIR Principles

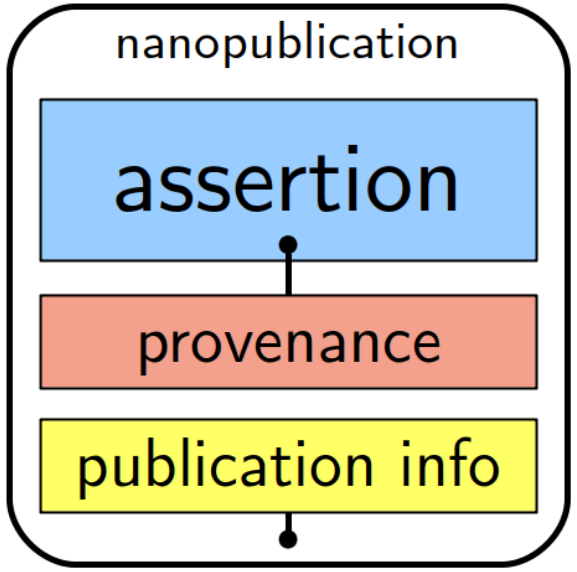
The FIP Ontology



Tobias Kuhn, Barbara Magagna & Erik Schultes

FAIR Enabling Resource (FER) types

FAIR Sub-principle	Type of FAIR Enabling Resource	Definition
F1	identifier service	A **service** that provides for any digital object (1) algorithms guaranteeing global uniqueness, (2) policy document that guarantees persistent and (3) resolution of the identifier to machine-actionable metadata describing the object and its location.
F2	metadata schema	A **specification** that defines metadata fields describing attributes of data or other digital objects.
F3	metadata-Data linking schema	A **specification** that provides a unique, persistent, (ideally) bi-directional, machine-actionable link between metadata and the data they describe.
F4	registry	A **service** that indexes metadata and data and provides search over that index.
A1.1	communication protocol	A **specification** that defines how messages are structured and exchanged.
A1.2	authorization and authentication service	A **service** that mediates access to digital objects according to specifed conditions.
A2	metadata preservation policy	A **document** that describes the conditions under which metadata are to be provisioned in the future (maybe part of a data management plan).
I1	knowledge representation language	A language **specification** whereby knowledge can be made processible by machines.
I2	structured vocabulary	A **specification** of uniquely identified and unambigous concepts with their definitions represented preferably using web standards.
I3	semantic model	A **specification** that defines qualified relations between entities describing data or other digital objects using structured vocabularies.
R1.1	usage license	A **document** that describes the conditions under which a digital object can be legally used.
R1.2	provenance model	A **specification** that defines metadata fields describing the origin and lineage of data or other digital objects.
R1.3	FIP as a whole	



SeaDataNet CDI metadata XML schema

Nanopublication

```
@prefix this: <http://purl.org/np/RAeS0-GmZr7-ysXVdVptQXT4IDRAt55i0dwNuiEQH2AgU> .
@prefix sub: <http://purl.org/np/RAeS0-GmZr7-ysXVdVptQXT4IDRAt55i0dwNuiEQH2AgU#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix np: <http://www.nanopub.org/nschema#> .
@prefix orcid: <https://orcid.org/> .
@prefix nt: <https://w3id.org/np/o/ntemplate/> .
@prefix npx: <http://purl.org/nanopub/x/> .
@prefix fip: <https://w3id.org/fair/fip/terms/> .

sub:Head {
  this: np:hasAssertion sub:assertion ;
    np:hasProvenance sub:provenance ;
    np:hasPublicationInfo sub:pubinfo ;
    a np:Nanopublication .
}

sub:assertion {
  sub:SeaDataNet_CDI_Schema a fip:Available-FAIR-Enabling-Resource , fip:FAIR-Enabling-Resource , fip:Metadata-schema ;
    rdfs:comment "XML schema for CDI metadata, expanded profile of ISO19139 https://www.wikidata.org/wiki/Q30079268" ;
    rdfs:label "SeaDataNet CDI metadata XML schema" ;
    rdfs:seeAlso <http://schemas.seadatanet.org/Standards-Software/Metadata-formats/SDN2_CDI_ISO19139_10.0.1.sch> .
}

sub:provenance {
  sub:assertion prov:wasAttributedTo orcid:0000-0001-9214-3217 .
}

sub:pubinfo {
  sub:sig npx:hasAlgorithm "RSA" ;
    npx:hasPublicKey
"MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQCD5O+7prx9soILN1GCLZfmTgs3LBza9g3HLUyCGVR/Y8abJQa4P6Z3py9i2L1qXcE7DTNBjaW8iJJIi6sQa/prx1pql
;
    npx:hasSignature
"AXMy3w+FTZ5EhaCK1BQ9QEbJ1cV88ULcgpN/+Jimdq5/YvPYXOaPTcQ+Z2Cqpy8o17iYbnXbdlNbo39Q8UibK7Fhpv/ORKv93UKTPezGRrSCvClbKX72YSITQcBwDbgbo
;
    npx:hasSignatureTarget this: .
  this: dct:created "2020-10-15T17:47:50.907+02:00"^^xsd:dateTime ;
    dct:creator orcid:0000-0001-9214-3217 ;
    npx:introduces sub:SeaDataNet_CDI_Schema ;
    nt:wasCreatedFromProvenanceTemplate <http://purl.org/np/RANwQa4ICWS5SOjw7gp99nBpXBasapwtZF1fIM3H2gYTM> ;
    nt:wasCreatedFromPubinfoTemplate <http://purl.org/np/RAA2MfqdBCzmz9yVWjKLXNbyfBNcwsMmOqcNUxkk1maIM> ;
    nt:wasCreatedFromTemplate <http://purl.org/np/RAHvHX5qjbdnYXsZWsRMO3KuFekGUFR6LuPjigZns9_VA> .
}
```

https://nanopub.org/wordpress/

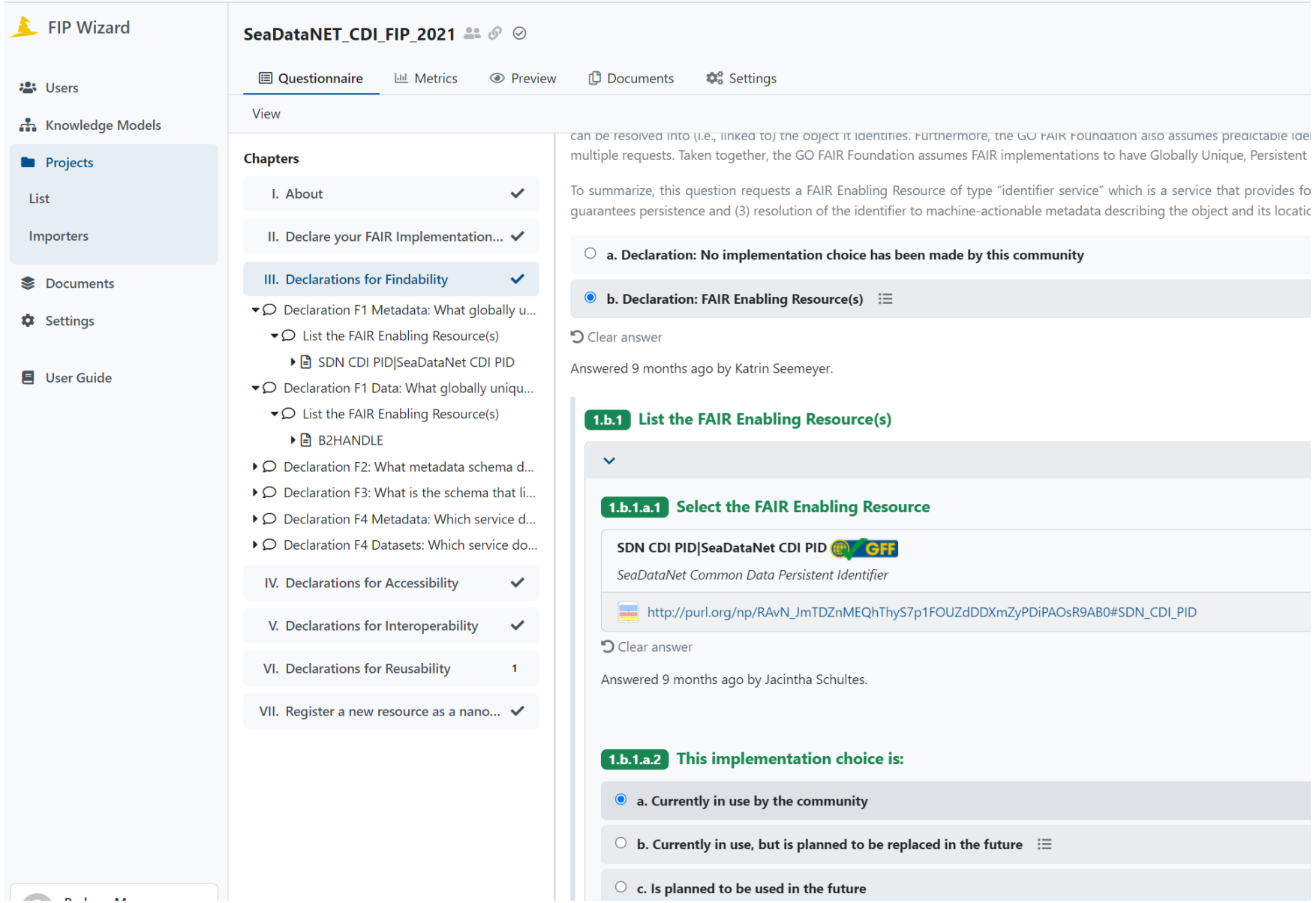
https://peerj.com/articles/cs-387/

Nanopub monitor: https://purl.org/nanopub/monitor

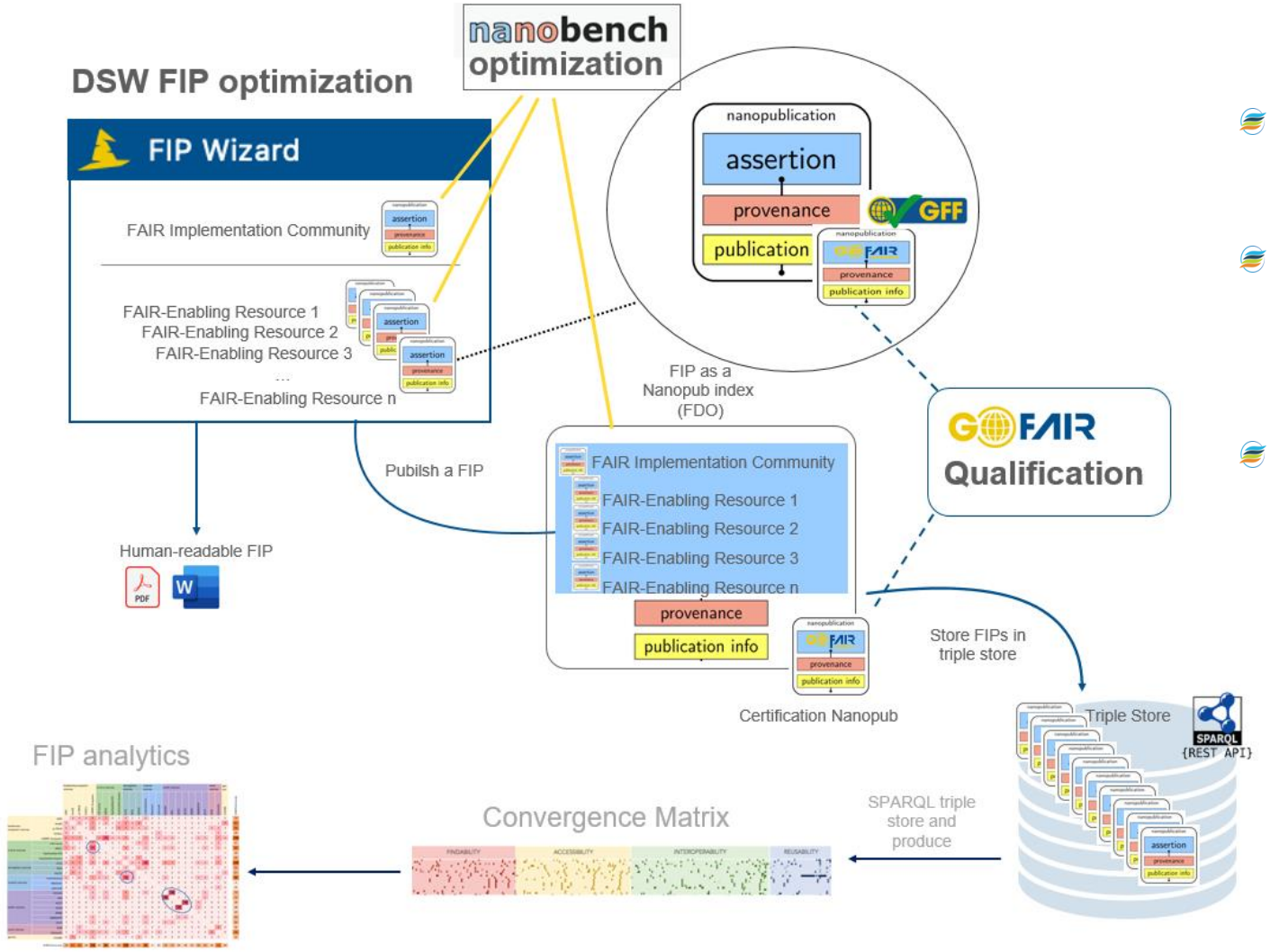
Nanopub server: https://np.petapico.org/

Nanobench: https://github.com/peta-pico/nanobench

The FIP Wizard



FIP workflow

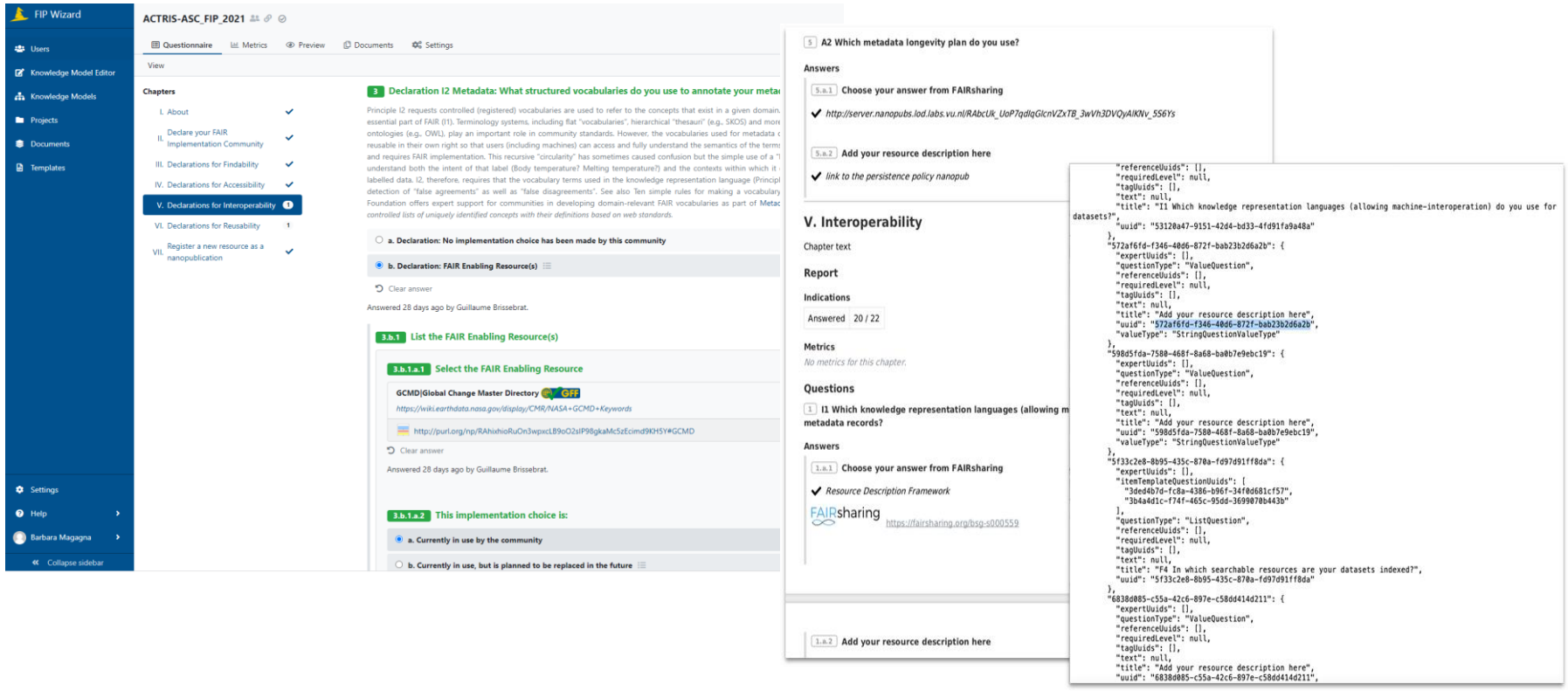


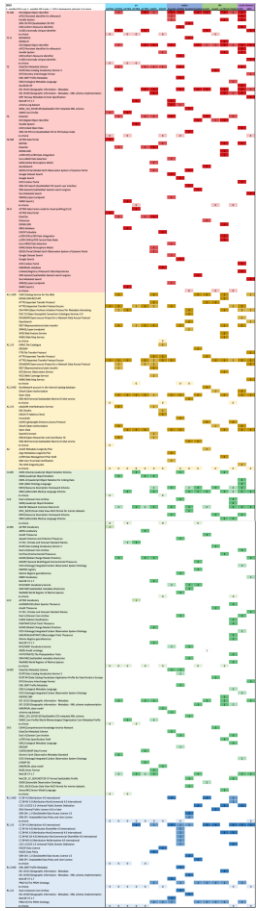
- Data Stewardship Wizard for questionnaire based assessments
- nanopublications for machine-readable statements about FAIR implementations
- a success story of collaboration between:

ENVRI FAIR GO FAIR CODEVENCE

ENVRI FAIR

FIP outputs





human friendly interface – human /machine readable outputs –FIP analysis – FAIR convergence

FAIR Convergence (FIP) Matrix

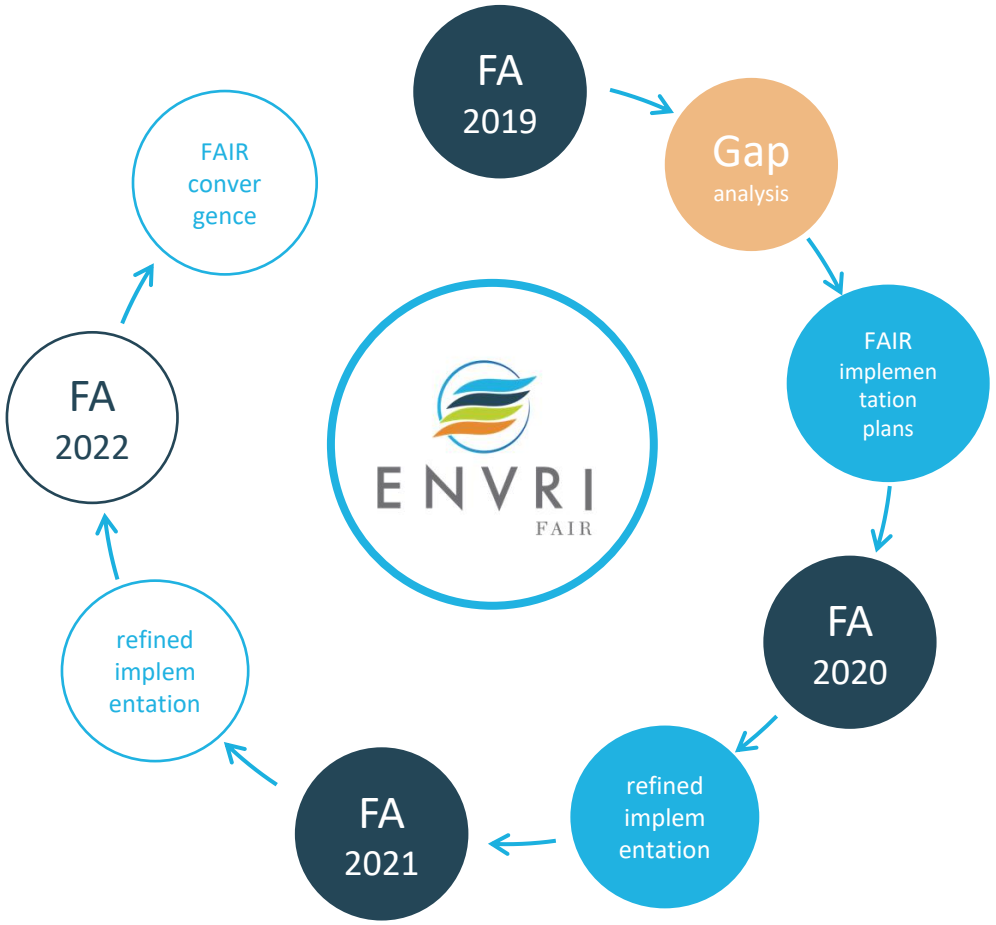
FAIR Enabling Resources

FAIR Communities

FIP

2021 3 - availabe FER in use, 2 - planned available FER, 1 - planned FER in development, 0 - no choice		air			water					earth	life						multi-domain	
		ACTRIS-ASC	IAGOS	EISCAT	ARGO	EMSO ERIC	LifeWatch marine	SDN CDI	SDNS SEXTANT	EPOS	AnaEE	AnaEE-Crea	Danubius	DiSSCo	eLTER-RI	LifeWatch Ecosystem	ICOS	SIOS
I3-D	CKAN|Comprehensive Knowledge Archive Network											3						
	DataCite Metadata Scheme				3													
	DCAT-AP|Data Catalog Vocabulary Application Profile for Data Portals in Europe												2					
	DwC-A|Darwin Core Archive						3									3		
	EISCAT Level 3 HDF5 archival format			3														
	eLTER Data Specification Draft														3			
	ERDDAP						2											
	Generic Earth Observation Metadata Standard																	
	HDF5 archival file format for data in Madrigal			3														
	I-ADOPT Framework														2			
	ICOS Ontology|Integrated Carbon Observation System Ontology																3	
	JCAMP-DX	3																
	MADRIGAL data model			3														
	NASA Ames Format		3															
	NetCDF CF-1.7	3	3						3				2				3	3
	NetCDF_CF_SDN|NETCDF CF format SeaDataNet Profile							3								3		
	OBOE|Extensible Observation Ontology										2							
	OceanSITES Data Format					3												
	ODV_ASCII|Ocean Data View ASCII format for marine datasets							3	3									
	openDS|Open Digital Specimens													1				
	SensorML|Sensor Model Language														3			
	WMO Core Profile|World Meteorological Organization Core Metadata Profile																2	
	no choice									0								
R1.1-MD	CC BY 4.0|Attribution 4.0 International				3	3		3	3	3	3		2	2		3		
	CC BY-NC 4.0|Attribution-NonCommercial 4.0 International															3		
	CC0 1.0|CC0 1.0 Universal Public Domain Dedication		2											2	3	3	3	
	GNU General Public License v3.0 or later											3						
	GPL-2.0-or-later|GNU General Public License v2.0 or later															3		

FIP process



- FAIR Assessments (FA) updated manually using FIPs
- Gap analysis
- FAIR Implementation Plans
- Refined implementations

t o w a r d s

FAIR Convergence

FAIR Convergence Matrix for ENVRI-FAIR

	2019	2020	2021
Different FERs used	132	145	152
By FAIR principle			
F1-MD	7	8	8
F1-D	8	8	7
F2	13	15	16
F3	6	6	6
F4-MD	17	20	26
F4-D	20	23	25
A1.1-MD	13	14	13
A1.1-D	11	11	12
A1.2-MD	5	10	11
A1.2-D	11	13	13
A2	5	6	7
I1-MD	7	7	7
I1-D	7	6	6
I2-MD	19	27	27
I2-D	18	20	20
I3-MD	15	17	20
I3-D	18	18	21
R1.1-MD	7	8	6
R1.1-D	11	11	8
R1.2-MD	6	6	7
R1.2-D	5	6	7

- The number of FERs increases…

FIP basic statistics

		2019	2020	2021
AIR	ACTRIS_DVAS	☑	☑	☑
	ACTRIS-Gres FIP	☑	☑	☑
	ACRIS-inSitu	☑	☑	☑
	ACTRIS_ARES	☐	☑	☑
	ACTRIS_CLU_FIP	☑	☑	☑
	ACTRIS-ASC	☑	☑	☑
	IAGOS	☑	☑	☑
	EISCAT_FIP	☑	☑	☑
WATER	ARGO	☑	☑	☑
	EMSO ERIC FIP	☐	☐	☑
	LW marine	☑	☑	☑
	SeaDataNet-CDI	☑	☑	☑
	SeaDataNet-Sextant	☑	☑	☑
LAND	EPOS	☐	☐	☑
LIFE	AnaEE	☑	☑	☑
	AnaEE-Crea	☑	☑	☑
	Danubius	☐	☐	☑
	DiSSCo_FIP	☐	☑	☑
	eLTER-RI	☑	☑	☑
	LWERIC Ecosystem	☑	☑	☑
multi-domain	ICOS FIP	☑	☑	☑
	SIOS FIP	☑	☑	☑
	Total count: 57	17	18	22

- Number of FIP Wizard questions: 21
- Average FIP length: 28 FERs
- Total numbers of FER declared: 193
- Total number of triple statements: 1852

FAIR Convergence

Resource overlap/FIP adoption

Convergence by …

Resource Quality

FAIRness/Interoperability

Resource Overlap

community vs community

2021	ACTRIS_DVAS	ACTRIS_GRES	ACTRIS_InSitu	ACTRIS_CLU	ACTRIS-ARES	ACTRIS_ASC	IAGOS	EISCAT	ArgoGdac2	EMSO	lw-marine	SeaDataNet-CDI	SeaDataNet-Sextant	EPOS-ERIC	Anaee	AnaEE_CREA	DANUBIUS	DiSSCo	eLTER-RI	LWERIC_Ecosystem	ICOS	SIOS	
ACTRIS_DVAS		11	11	8	8	11	9	2	6	4	4	2	5	3	7	4	4	3	4	5	7	7	125
ACTRIS_GRES	11		13	13	11	18	19	4	10	5	9	3	9	7	9	5	7	5	9	7	16	13	203
ACTRIS_InSitu	11	13		10	8	12	13	3	8	5	5	5	8	4	7	4	5	3	6	7	11	13	161
ACTRIS-ARES	8	11	8	8		9	12	3	7	4	5	1	4	5	4	4	5	2	5	5	13	7	130
ACTRIS_CLU	8	13	10		8	12	10	3	8	4	6	2	7	6	5	4	5	5	7	6	10	7	146
ACTRIS_ASC	11	18	12	12	9		14	3	9	5	8	2	8	5	8	4	6	5	7	6	10	11	173
IAGOS	9	19	13	10	12	14		4	11	6	10	7	12	7	9	5	7	6	12	10	20	13	216
EISCAT	2	4	3	3	3	3	4		5	2	3	3	3	2	2	2	1	1	5	1	7	2	61
ArgoGdac	6	10	8	8	7	9	11	5		8	7	6	10	3	6	5	6	4	5	7	14	6	151
EMSO	4	5	5	4	4	5	6	2	8		4	4	7	3	4	4	4	1	2	6	7	4	93
lw-marine	4	9	5	6	5	8	10	3	7	4		3	7	4	6	5	5	9	10	9	13	5	137
SeaDataNet-CDI	2	3	5	2	1	2	7	3	6	4	3		10	2	3	2	4	1	4	7	8	6	85
SeaDataNet-Sextant	5	9	8	7	4	8	12	3	10	7	7	10		4	8	5	6	3	8	9	12	9	154
EPOS-ERIC	3	7	4	6	5	5	7	2	3	3	4	2	4		3	3	1	4	7	4	8	5	90
Anaee	7	9	7	5	4	8	9	2	6	4	6	3	8	3		7	6	2	6	7	8	8	125
AnaEE_CREA	4	5	4	4	4	4	5	2	5	4	5	2	5	3	7		4	2	3	5	7	3	87
DANUBIUS	4	7	5	5	5	6	7	1	6	4	5	4	6	1	6	4		1	3	5	8	6	99
DiSSCo	3	5	3	5	2	5	6	1	4	1	9	1	3	4	2	2	1		5	3	7	2	74
eLTER-RI	4	9	6	7	5	7	12	5	5	2	10	4	8	7	6	3	3	5		6	12	8	134
LWERIC_Ecosystem	5	7	7	6	5	6	10	1	7	6	9	7	9	4	7	5	5	3	6		9	6	130
ICOS	7	16	11	10	13	10	20	7	14	7	13	8	12	8	8	7	8	7	12	9		11	218
SIOS	7	13	13	7	7	11	13	2	6	4	5	6	9	5	8	3	6	2	8	6	11		152
	125	203	161	146	130	173	216	61	151	93	137	85	154	90	125	87	99	74	134	130	218	152	2944

What are the benfits for ENVRI-FAIR?

- Increase our understanding of the FAIR principles and their advantages for the Ris
- Compile a technology landscape of the RIs
- Assess the implementation status for each RI (= their usage status of FERs)
- Detect information and implementation gaps
- Discover strengths
- Compare implementations by different RIs
- Evaluate possible technology take-ups for improvements
- Prioritize FAIR improvements
- Include chosen FAIR improvements in RI implementation plans

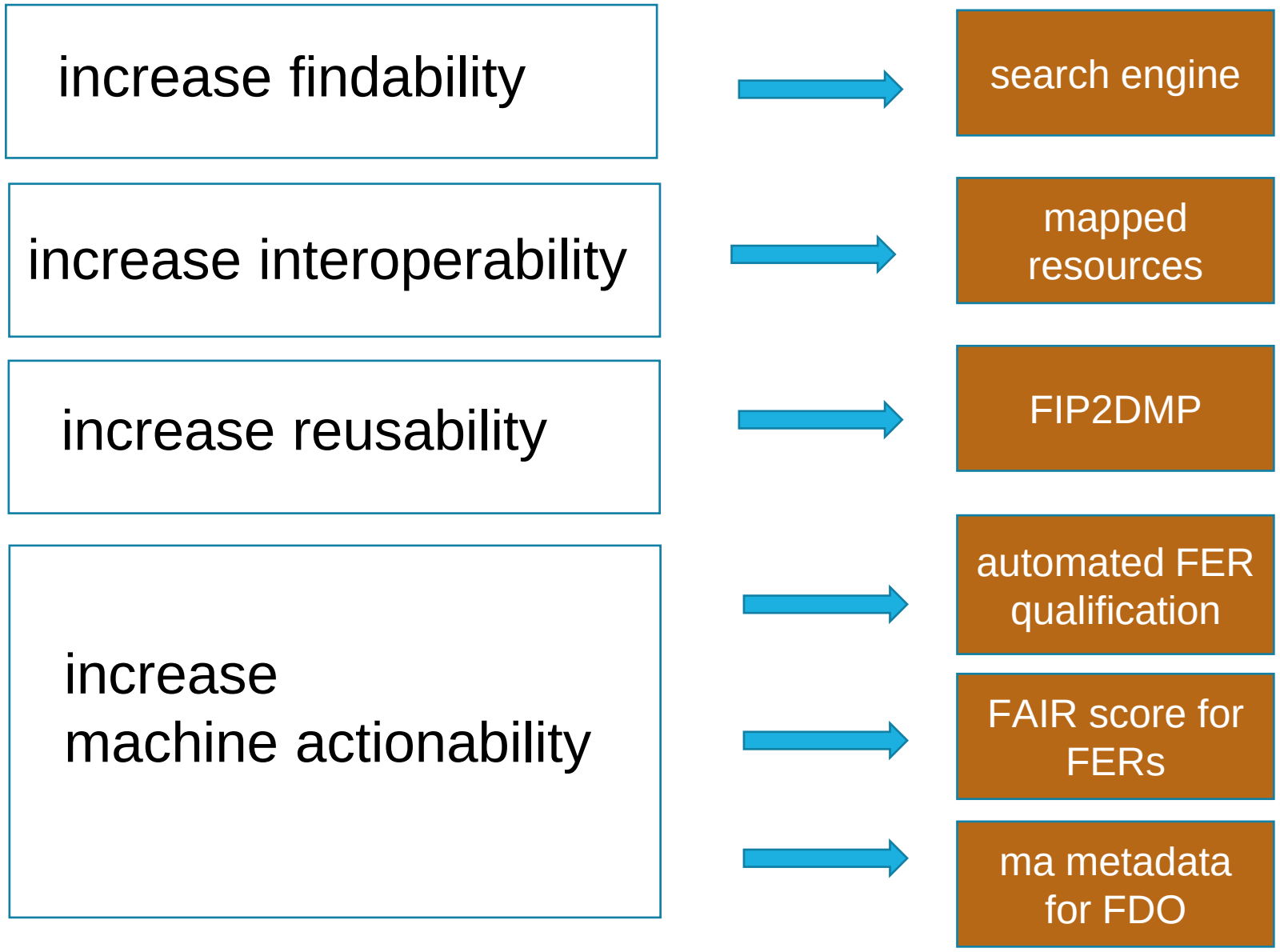
Limits of FIPs in ENVRI-FAIR

- Challenge to address the right accuracy level of a FER
- Challenge to define a community… in ENVRI-FAIR the community is built around a repository
- Convergence might result but is difficult to demonstrate
  - Very mature participating RIs (on the ESFRI level), most choices were made before the beginning of the project
- Interoperability is often achieved by providing mappings into other standards. This is not yet covered in the FIPs.

Who is using FIPs?

- CODATA/GO FAIR Pre-Symposium workshops 2020 (25 FIPs)
- ENVRI-FAIR (14 RIs, 56 FIPs)
- VODAN Africa
- Health-Holland projects (TWOC and C4Yourself)
- ODISSEI
- EOSC Nordic
- Dutch Cancer Foundation
- WorldFAIR
- NIAD/NIH office of data science strategy

Towards a professional FIP service



Background and Links

- Magagna B, Schultes EA, Suchánek M, Kuhn T (2022) FIPs and Practice. https://doi.org/10.3897/rio.8.e94451
- Schultes E., Magagna B., Hettne K.M., Pergl R., Suchánek M., Kuhn T. (2020) Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. https://doi.org/10.1007/978-3-030-65847-2_13
- Schultes EA, Gregory A, Magagna B (2022) Emerging FAIR Ecosystem(s): A Practical Perspective. https://doi.org/10.3897/rio.8.e94149
- Schultes EA, Magagna B, Kuhn T, Suchánek M, Bonino da Silva Santos LO, Mons B (2022) The Comparative Anatomy of Nanopublications and FAIR Digital Objects. https://doi.org/10.3897/rio.8.e94150
- H.P. Sustkova, K.M. Hettne, P. Wittenburg, A. Jacobsen, T. Kuhn, R. Pergl,… & E. Schultes. FAIR convergence matrix: Optimizing the reuse of existing FAIR-related resources. doi: 10.1162/dint_a_00038
- Making the ENVRIs „FIPs for purpose“ through FAIR Convergence workshop (2022) archive: https://osf.io/7n5yp/
- FIP Ontology
- FIP Wizard Tool https://fip-wizard.ds-wizard.org/
- Nanopublications and NanoBench by Tobias Kuhn on GitHub
- SPARQL query for making the FAIR matrix
- FAIR Digital Object Forum WG FDO FIPP