FAIR Implementation Profile

Erik Schultes

30 January 2025

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Finanziato dall'Unione europea NextGenerationEU
Ministero dell’Università e della Ricerca
Italiadomani

Agenda

13:00-13.30 (30 minutes)	Welcome - Recap (Barbara)
13:30-13:50 (20 minutes)	What is convergence and why it is important (Erik)
13:50-14:00 (10 minutes)	FIP analysis: the FIP matrix (Erik)
14:00-14:20 (20 minutes)	**Breakout: Explore the FIP matrix** https://fip-matrix.streamlit.app
14:20-14:35 (15 minutes)	**Break**
14:35-14:55 (20 minutes)	FIP analysis: statistical features (Erik)
14:55- 15:55 (60 minutes)	FIP analysis: graph visualization (Kristina)
15:55-16:05 (10 minutes)	**Break**
16:05-16:30 (25 minutes)	How to become an implementer - next steps (Barbara)
16:30-16:50 (20 minutes)	**Exam**
16:50-17:00 (10 minutes)	Feedback

OSF Project Space
Common notes

What is FAIR Convergence?

13:30-13:50
(Erik)

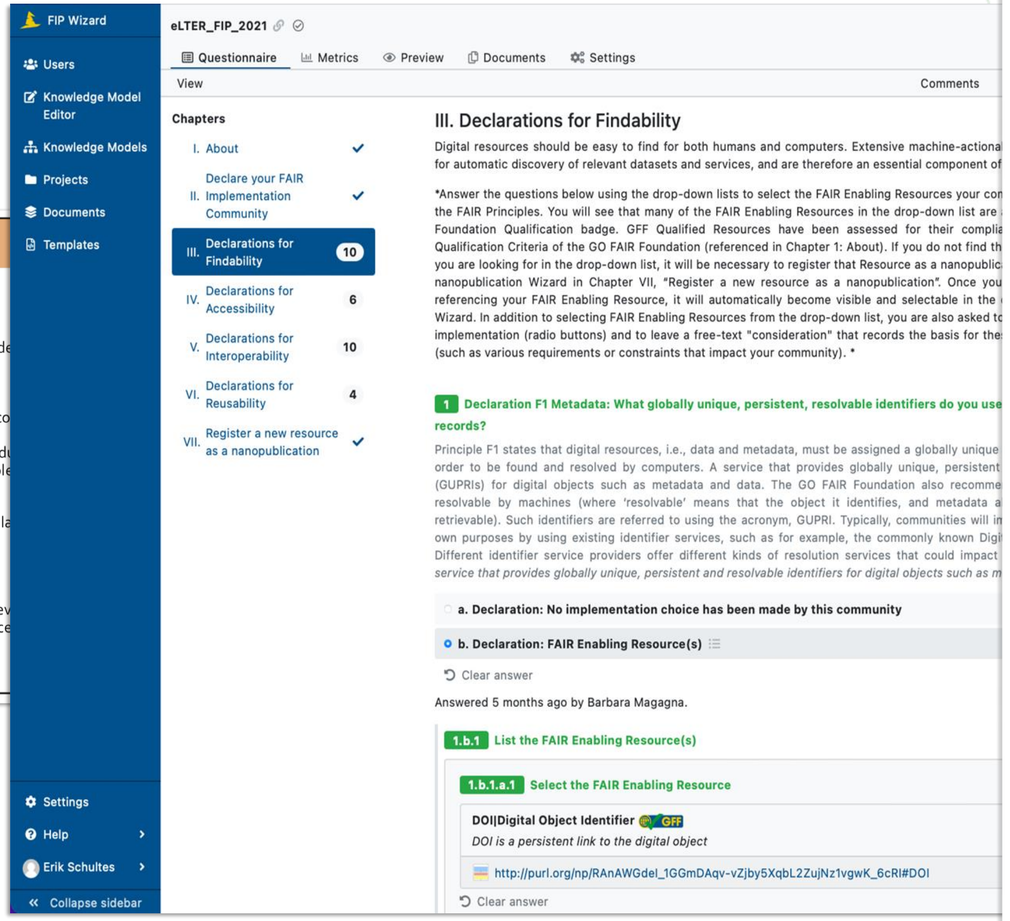
Box 2 | The FAIR Guiding Principles

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 de
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 co
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedu
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 la
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 relev
R1.1. (meta)data are released with a clear and accessible data usage lice
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards



eLTER_FIP_2021 *(Nanopub Index)*

[home]

This:

get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAxmL-ZZYM9pqWK-wWGEiKzHUnRLWqZ0yW9tQ1PzfwXh0

Creation Time:

March 10, 2022 10:26:13 AM UTC

Creators:

- https://fip-wizard.ds-wizard.org

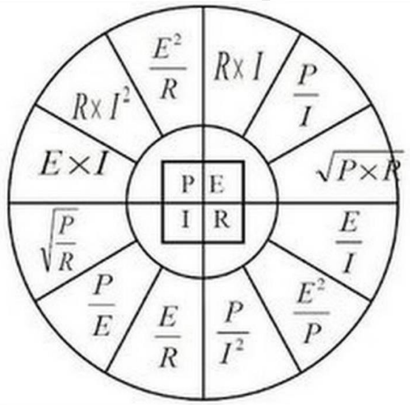
Derived From:

- https://w3id.org/fip/wizard/7095bdd2-3b99-4abd-a526-a9a9e7ae47e9

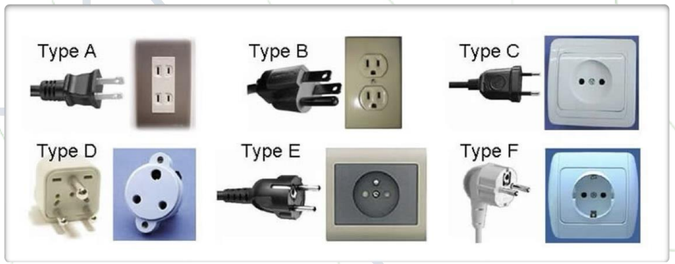
Includes as Elements:

get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAbe-1iuraXDmV9FEdosOdd-CNr3QiAj_hZo_E2i8Lge4
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAkag-9797lsc2qqMVyc1At_eypQwEWM-2hqwfsoUoba8
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAoVxAm0B1ZPTm2TuZAHhpOo4WNJHpMNooDKC-_lkUS5g
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAEF6ozjqgvdwmh94c2UIf3AJP5ez63yWF63IRY9z5cIk
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAhN3WHzb77s_XIvYN5NBXbnWWxJdECSFv33zzERlyi8Q
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAMGmK-RtiPQUi0UsdfD60be3fmWZ2K7h3JfivCmzXZ1g
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAGPbavsobuHn-xyZwH9daBDjTDmKNn-f6dtqQTsSSHf0
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAoD8rctuAeX8wCdFg8nwyPnWk4AiGsqzOoSdJBHHDQeA
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RA6ZlPmbIoocTOF0cKz5eICH7BrhVJujf9HhjdWLd3j3w
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAdYQPj2rPdEfXSzMLWrbBsABbFfoV_h17BtV7jadrvUw
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RA0sqstdVSAbAl4flpc7zu8L-PmPyxocpUTpbs5kkyKK8
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RA5ZU134suAOPOw-yC5BDaTNu3e_WlTn04tLbYi3ZfShg
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAQIkRqjG_WDHUfZ0cOAaE71cAtz8ZjGnS7-xsNGSVZ_I
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAme447stXVxo4VmTPa1QlWHiP-ElmMui3__aAUlmUW9I
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RApvPwCokluoVn6D4zU3Bak0bWvHs1eXayd14V0Bz1eSc
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAZCYIfDQSB3R_Hjw4TOP1us3vjcpyxlVq87CypaH-SyE
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RA9vOV8ITTG3_5kywrtFgFx_qWNx8E0S5I3FI_Arnh-nc
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAOtYhHpJH-vsiis2jmBHbuXzSUOzAQEwi_58cHlDDFQU
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAs3Fzelnw96XFcCsbNM4_EUKFkhFKwhKSlQB3aLk83T8
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAgPyeIxelHPdR1f6cwBkOByP2t2WtMyzZ3AlIwfVEoY0
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAJD83HyflH-AuGz8mWCCulvKN2ifAAeu7CxGnRrvYKfI
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAcYXIyheegOosSeMJ45fxYeHQtyyyU_japTurslN5vDY
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RA5ZYCM2QjtqBcTNXk-TZKMBYN5aeDCTUx1k2F8_uHU2k
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAZhqmKtJ2W34pb1Wbi2hx3xY2WCV8fmPtTIbvaWlYyjI
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RA6FRBAe7xEaicDBw-fRIodCn6BU4qeUNi0lbvmGK0qNk
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAzPpcMpfdsZJAFFqLHXudlTRPXhxe9X48gDmxawX4wY4
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAV1mqQO8wKZNxyEEgoBBO4oXZDdmhUBz7BCGYybu7PjY
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAnLDJOlaR3isI6vMrRIzs_uzSoSUxmsjm2o7msGyz66s
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAay9n1IqFC4Te1l9f1XlYgS1OCt_YV2l5v2oePkIFqWY
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RACeiP-vdd84N7sMZFtACWYGPRdAdkESalAjDzRFAA_vI
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAweNCC1xuU6FmQF1CXjVdbIhybMquPGXkRTPOetGwBY
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAwxuE0D8ZVnNCQUZk8ay_9Emu28ARdYOn8aVeYm33Hps
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RABESomxrqh8fsmT3gJbsFmlyU7jAy8coqHGR7pbv0X7g
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAQqj0ni6zsbdJDAjGlQeYu602h-7Jn26z9IHB8zZUwyI
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RARDZlbbkdEkLD2qJK0HHC8zXRvz42qRwkssNe0FlD1FI
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RATWgma0_v5l_hfbI3tB1vJgX3Q3BakkuNvGfC8GYC8MM
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAvqckES7ZKj-ttIY2ozAlgWcLOEwIJT8rf_QDbmELI4w
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RA4WwQ_a0P_7PBxa6FHkkC5d5pQ7HrCrPZLEQO_AAR9O0
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAGEvNpUGfDPPQIp3Lyfp021GXLAwvxaDepMb6TNy_9lc
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAucVpB07i4FlbOFXzJy8RM6wiB6zUGuPOK_3nN-GIvOA
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RA03pr4PhIgwip56qXhIdCmRdpcjr6W4kD8AApTi8tE-w
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAE_30u4ksHYOpLyWaxCc71h5LK-KbkfUKNQUV1j-aFhg
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAnMg6XriX8wF9D8EnKyuHpbG_yopi0yjK_WxW5VyRwuU
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RAKRuT-KuI5J1nClE5nc9iQ4KYlFzlUyaLZ2rpfBoPY0s
get (trig, nq, xml, jsonld, trig.txt, nq.txt, xml.txt, jsonld.txt) http://purl.org/np/RA1Yh1zHJDgAltaNVZmWvdtFf0A4joxSn2Ggku8byQ31w

FAIR Principles



GO FAIR



FAIR Implementations

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

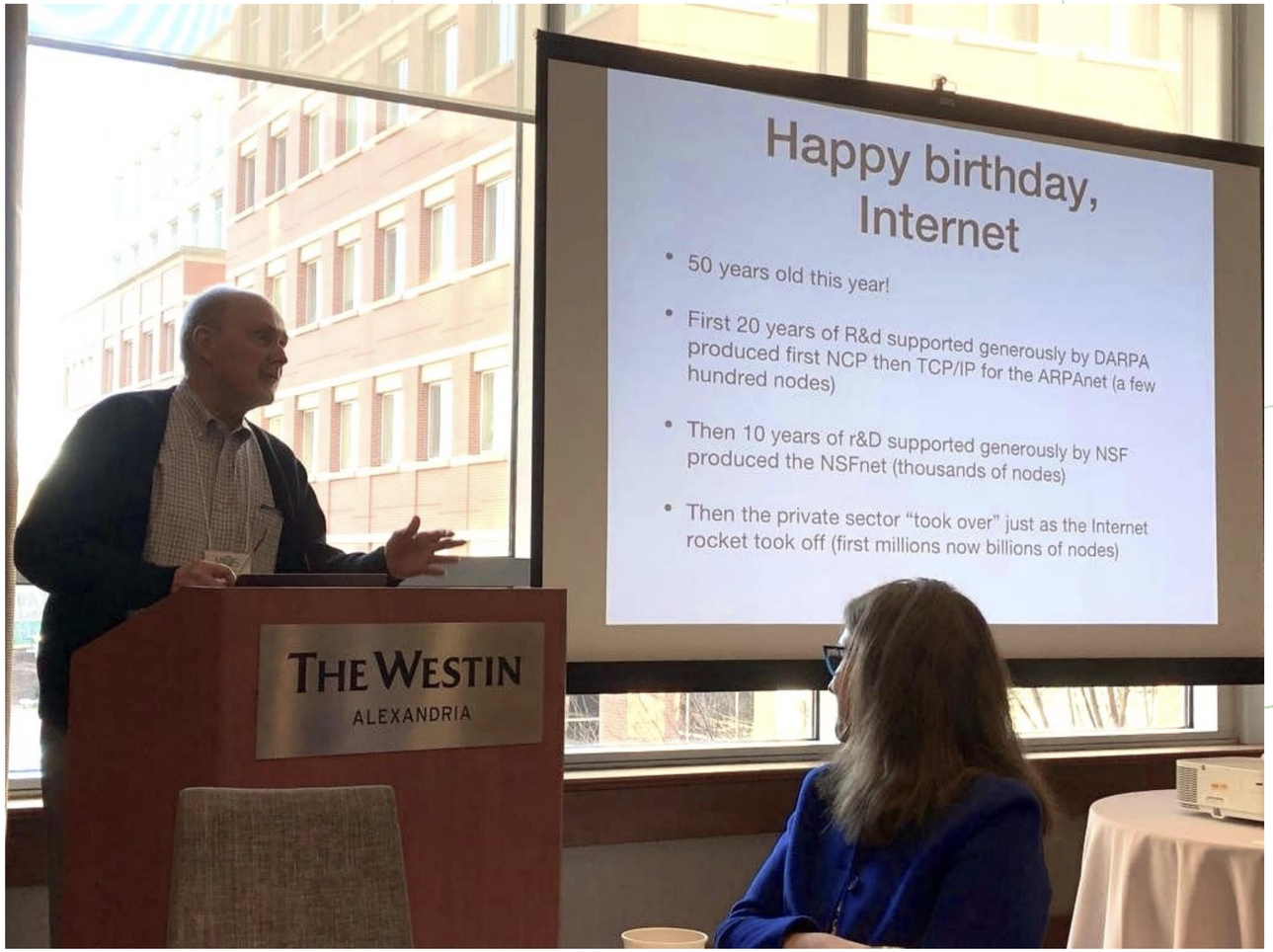
Societies have seen large infrastructures emerge when new technologies become available. From history we see that such infrastructures can have a huge influence on all aspects of societal life. Moreover, some patterns appear to reoccur in the evolution of such infrastructures. *Early visions* about the 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.

[1] *Creolization* is a term used to describe the development of culture and languages. It describes a process in which continuously new cultures/languages emerge and are mixing resulting in a broad spectrum of them.
[2] http://rd-alliance.org
[3] http://www.c2camp.org



https://b2share.eudat.eu/records/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*.

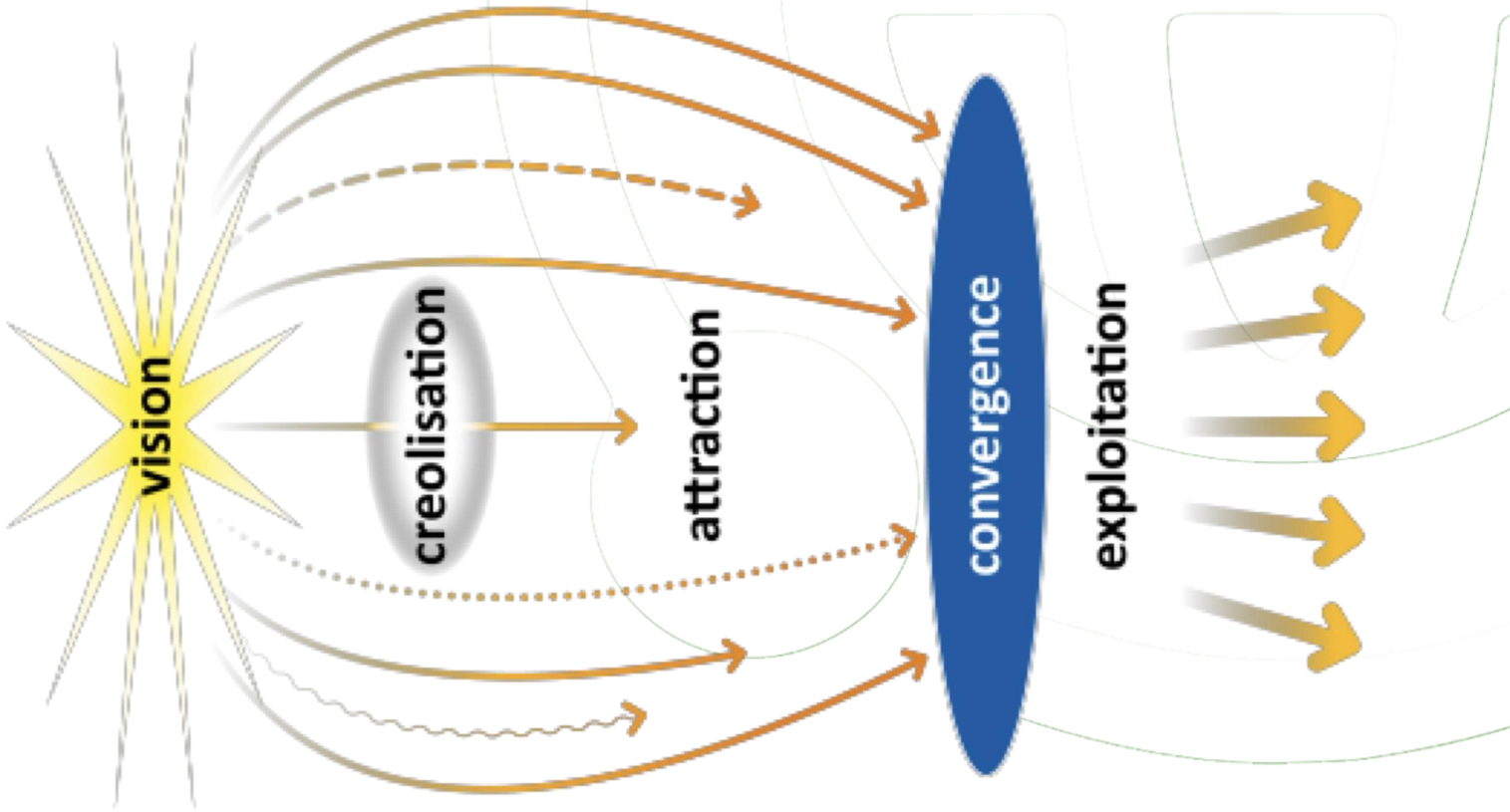
It appears that the "data infrastructure" is evolving into such a large infrastructure, with a potentially large influence on societies, industry and science. In order to gain new insights about complex relationships in nature, societies and minds, by integrating data from different silos we have seen 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.

---

[1] *Creolization* is a term used to describe the development of culture and languages. It describes a process in which continuously new cultures/languages emerge and are mixing resulting in a broad spectrum of them.
[2] http://rd-alliance.org
[3] http://www.c2camp.org



https://b2share.eudat.eu/records/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

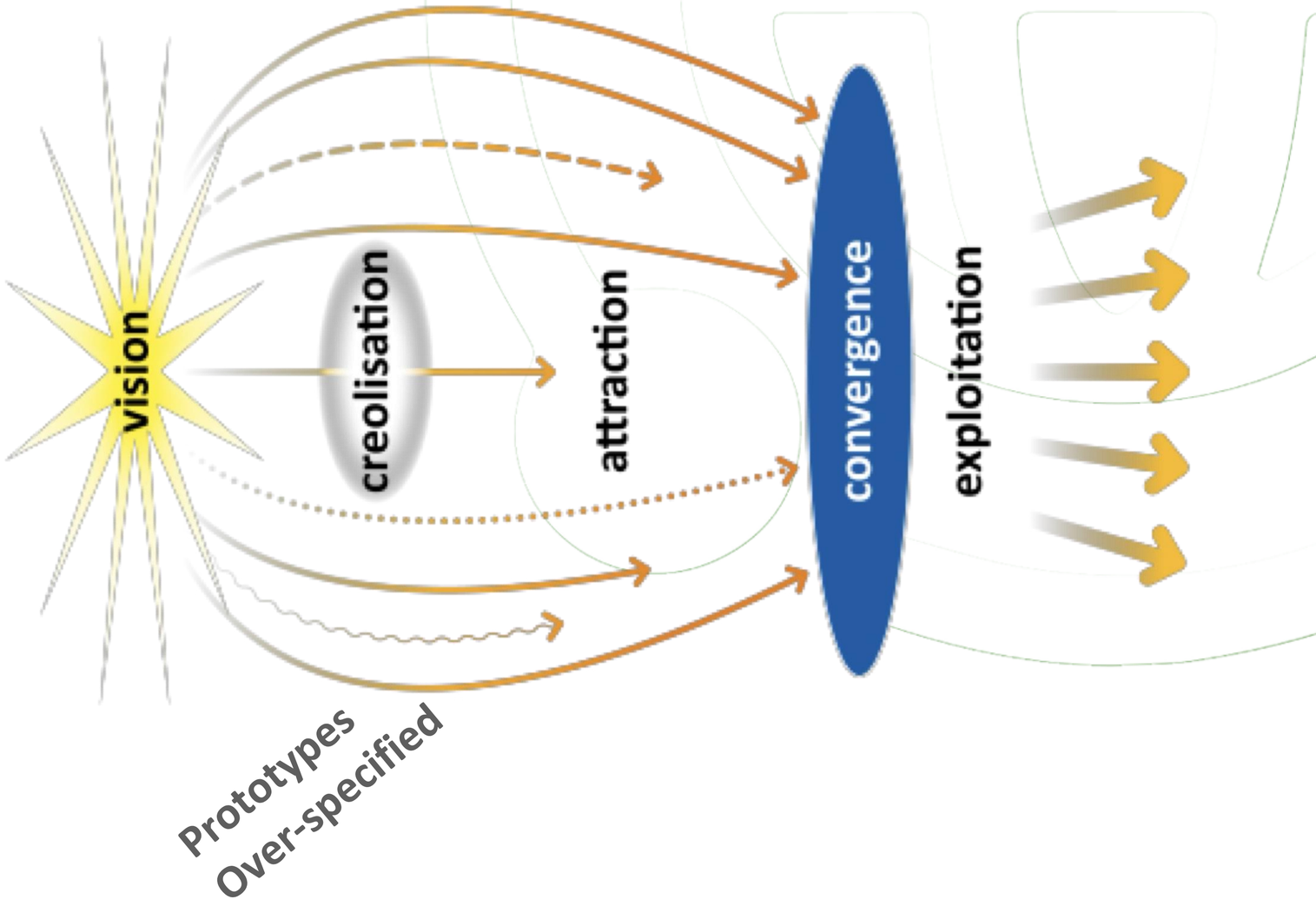
Societies have seen large infrastructures emerge when new technologies become available. From history we see that such infrastructures can have a huge influence on all aspects of societal life. Moreover, some patterns appear to reoccur in the evolution of such infrastructures. *Early visions* about the 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.

[1] *Creolization* is a term used to describe the development of culture and languages. It describes a process in which continuously new cultures/languages emerge and are mixing resulting in a broad spectrum of them.
[2] http://rd-alliance.org
[3] http://www.c2camp.org



https://b2share.eudat.eu/records/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

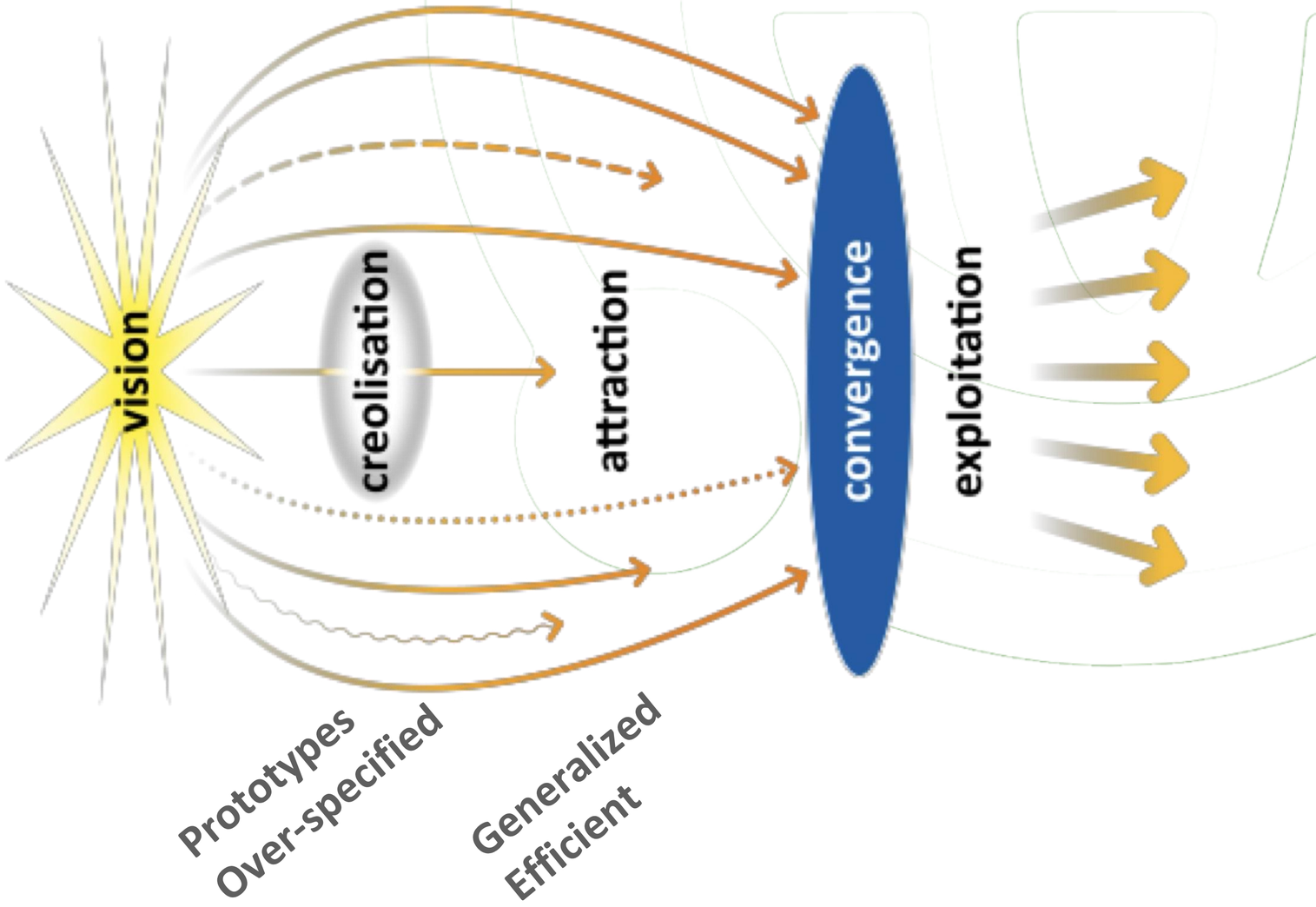
Societies have seen large infrastructures emerge when new technologies become available. From history we see that such infrastructures can have a huge influence on all aspects of societal life. Moreover, some patterns appear to reoccur in the evolution of such infrastructures. *Early visions* about the 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.

[1] *Creolization* is a term used to describe the development of culture and languages. It describes a process in which continuously new cultures/languages emerge and are mixing resulting in a broad spectrum of them.
[2] http://rd-alliance.org
[3] http://www.c2camp.org



https://b2share.eudat.eu/records/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

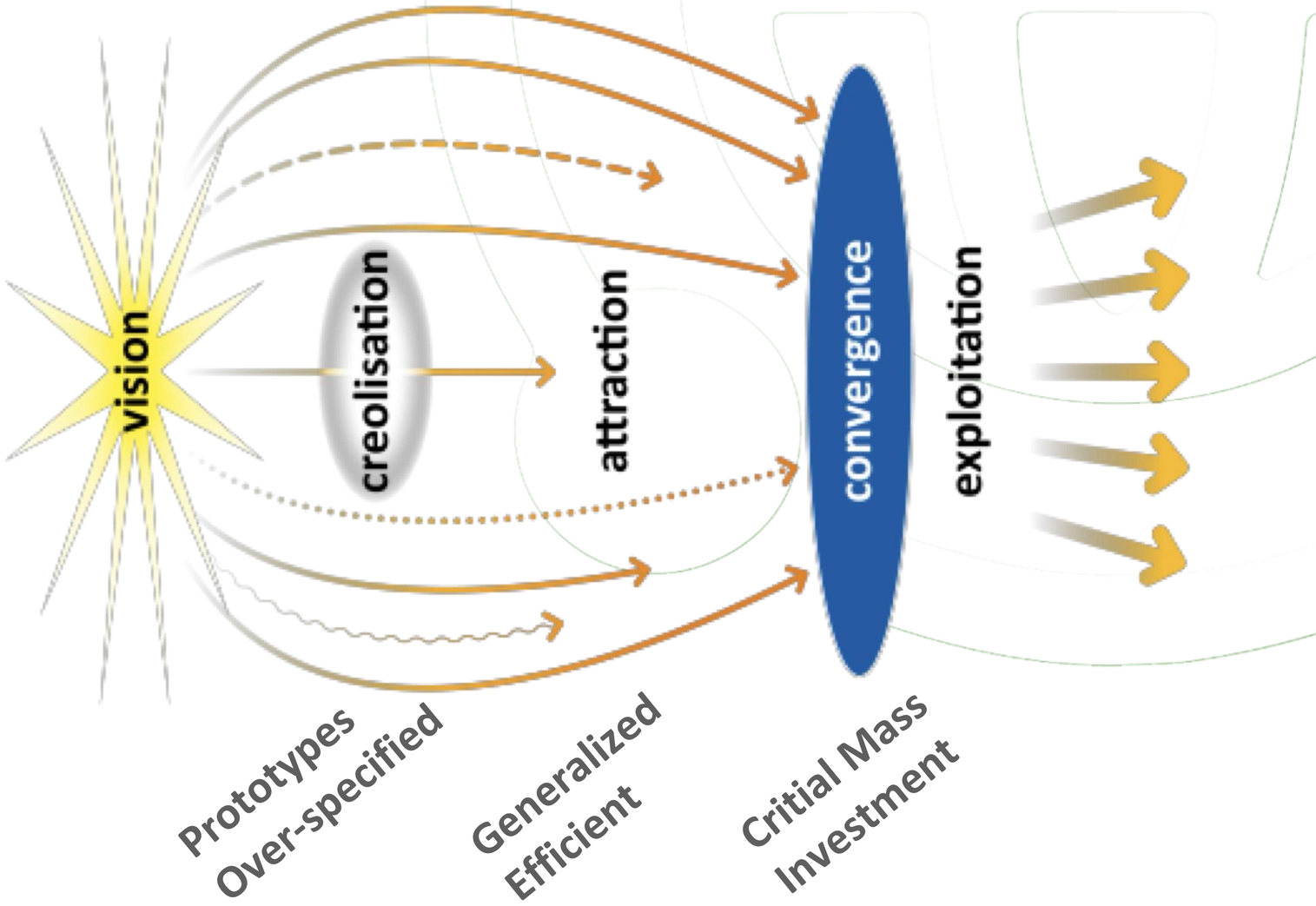
Societies have seen large infrastructures emerge when new technologies become available. From history we see that such infrastructures can have a huge influence on all aspects of societal life. Moreover, some patterns appear to reoccur in the evolution of such infrastructures. *Early visions* about the 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.

[1] *Creolization* is a term used to describe the development of culture and languages. It describes a process in which continuously new cultures/languages emerge and are mixing resulting in a broad spectrum of them.
[2] http://rd-alliance.org
[3] http://www.c2camp.org



https://b2share.eudat.eu/records/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

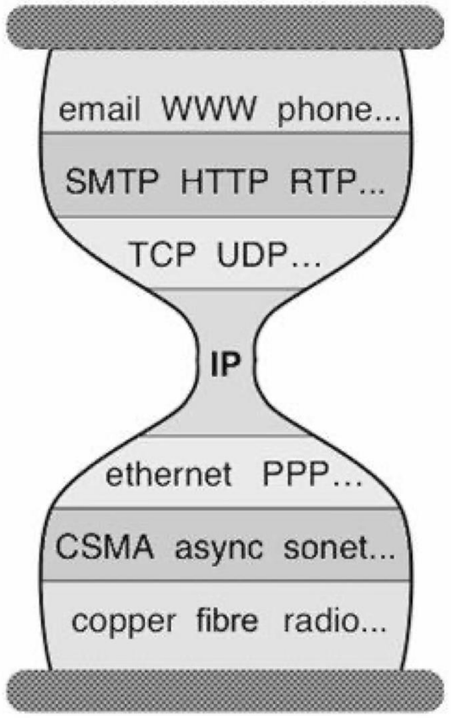
Societies have seen large infrastructures emerge when new technologies become available. From history we see that such infrastructures can have a huge influence on all aspects of societal life. Moreover, some patterns appear to reoccur in the evolution of such infrastructures. *Early visions* about the 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.

[1] *Creolization* is a term used to describe the development of culture and languages. It describes a process in which continuously new cultures/languages emerge and are mixing resulting in a broad spectrum of them.
[2] http://rd-alliance.org
[3] http://www.c2camp.org



Application

Implementation

https://b2share.eudat.eu/records/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

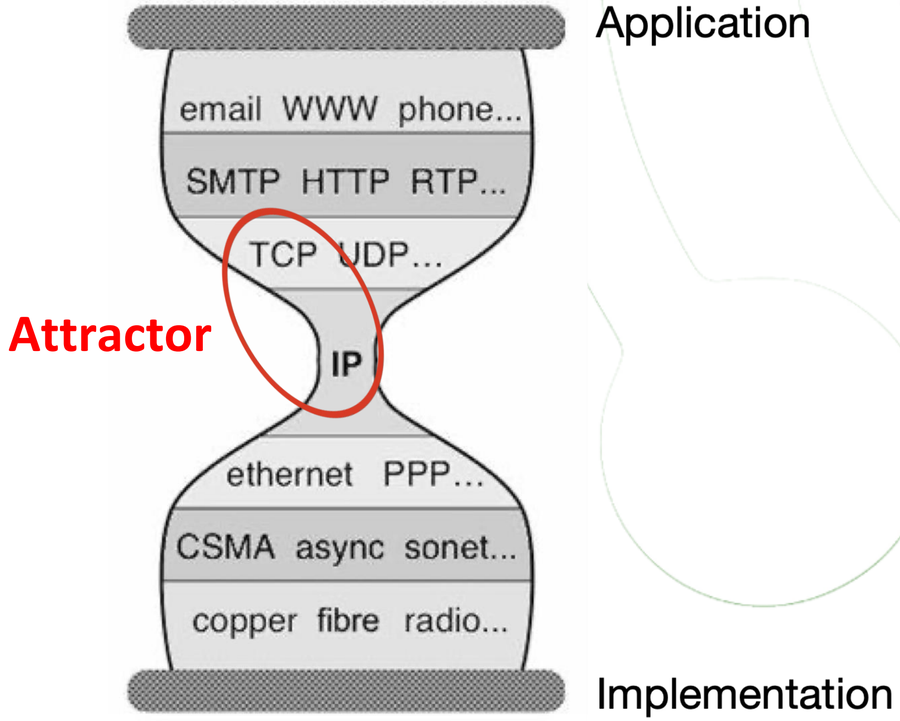
Societies have seen large infrastructures emerge when new technologies become available. From history we see that such infrastructures can have a huge influence on all aspects of societal life. Moreover, some patterns appear to reoccur in the evolution of such infrastructures. *Early visions* about the 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.

[1] *Creolization* is a term used to describe the development of culture and languages. It describes a process in which continuously new cultures/languages emerge and are mixing resulting in a broad spectrum of them.
[2] http://rd-alliance.org
[3] http://www.c2camp.org



https://b2share.eudat.eu/records/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

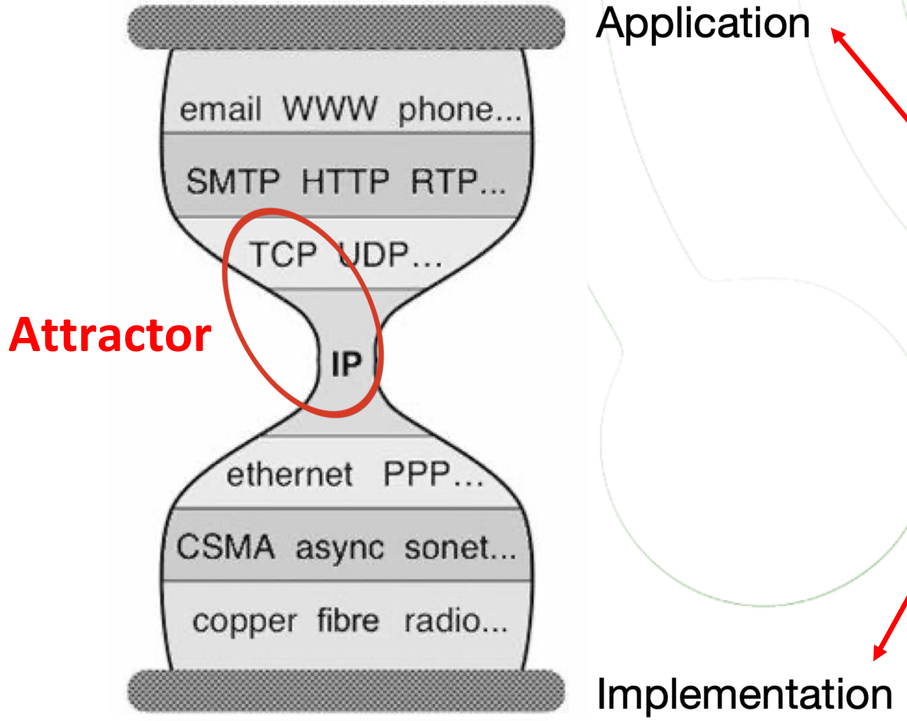
Societies have seen large infrastructures emerge when new technologies become available. From history we see that such infrastructures can have a huge influence on all aspects of societal life. Moreover, some patterns appear to reoccur in the evolution of such infrastructures. *Early visions* about the 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.

[1] *Creolization* is a term used to describe the development of culture and languages. It describes a process in which continuously new cultures/languages emerge and are mixing resulting in a broad spectrum of them.
[2] http://rd-alliance.org
[3] http://www.c2camp.org



https://b2share.eudat.eu/records/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

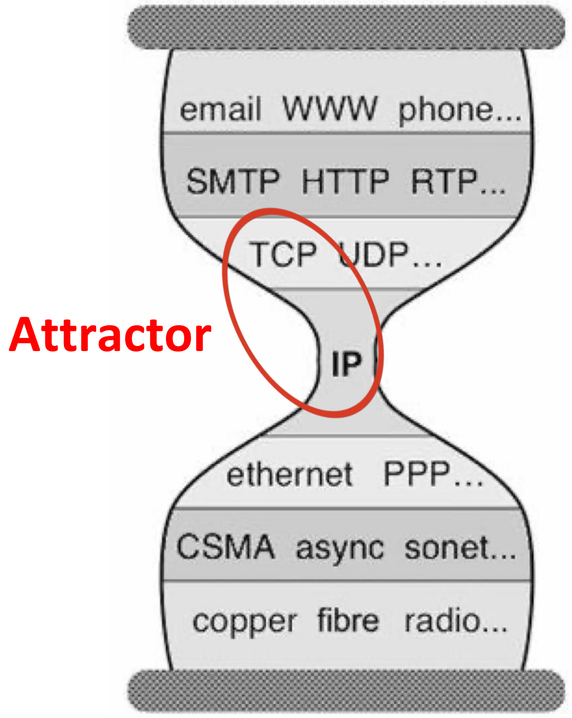
Societies have seen large infrastructures emerge when new technologies become available. From history we see that such infrastructures can have a huge influence on all aspects of societal life. Moreover, some patterns appear to reoccur in the evolution of such infrastructures. *Early visions* about the 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.

[1] *Creolization* is a term used to describe the development of culture and languages. It describes a process in which continuously new cultures/languages emerge and are mixing resulting in a broad spectrum of them.
[2] http://rd-alliance.org
[3] http://www.c2camp.org



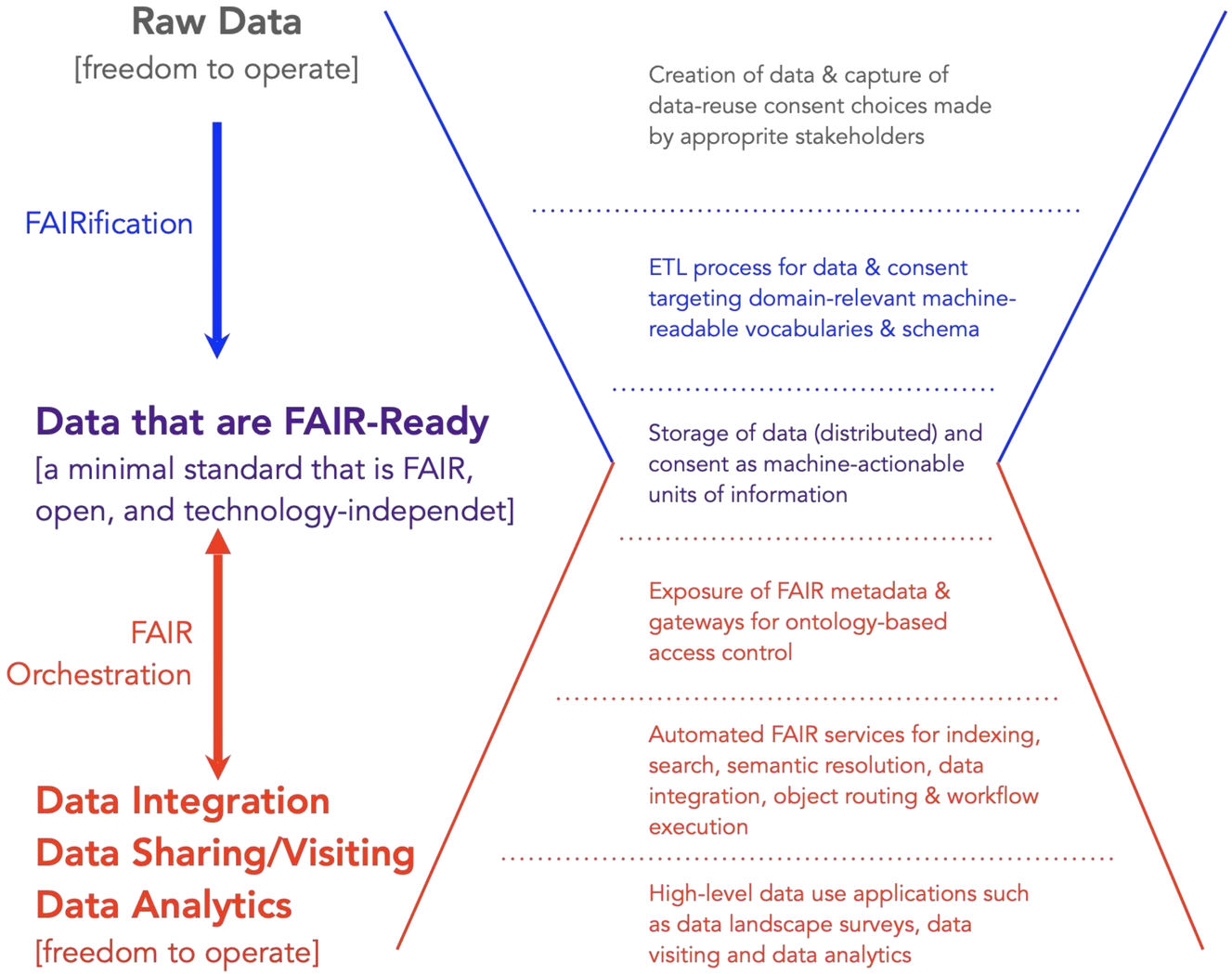
Attractor

Maximum freedom to operate

- Minimal standard
- Rough consensus, running code
- Voluntary participation
- Critical mass of users

https://b2share.eudat.eu/records/4e8ac36c0dd343da81fd9e83e72805a0

FAIR Hourglass



Schultes, Erik. 'The FAIR Hourglass: A Framework for FAIR Implementation'. 1 Jan. 2023 : 13 – 17.

https://content.iospress.com/articles/fair-connect/fc221514

FAIR Hourglass

Raw Data
[freedom to operate]

FAIRification

Data that are FAIR-Ready
[a minimal standard that is FAIR, open, and technology-independet]

FAIR Orchestration

Data Integration
Data Sharing/Visiting
Data Analytics
[freedom to operate]

Creation of data & capture of data-reuse consent choices made by approprite stakeholders

ETL process for data & consent targeting domain-relevant machine-readable vocabularies & schema

Storage of data (distributed) and consent as machine-actionable units of information

Exposure of FAIR metadata & gateways for ontology-based access control

Automated FAIR services for indexing, search, semantic resolution, data integration, object routing & workflow execution

High-level data use applications such as data landscape surveys, data visiting and data analytics

Schultes, Erik. 'The FAIR Hourglass: A Framework for FAIR Implementation'. 1 Jan. 2023 : 13 – 17.
https://content.iospress.com/articles/fair-connect/fc221514

FAIR Hourglass

Raw Data
[freedom to operate]

FAIRification

Data that are FAIR-Ready
[a minimal standard that is FAIR, open, and technology-independet]

FAIR Orchestration

Data Integration
Data Sharing/Visiting
Data Analytics
[freedom to operate]

Creation of data & capture of data-reuse consent choices made by approprite stakeholders

ETL process for data & consent targeting domain-relevant machine-readable vocabularies & schema

Storage of data (distributed) and consent as machine-actionable units of information

Exposure of FAIR metadata & gateways for ontology-based access control

Automated FAIR services for indexing, search, semantic resolution, data integration, object routing & workflow execution

High-level data use applications such as data landscape surveys, data visiting and data analytics

Wizard CEDAR Workbench

nanobench BioPortal

R jupyter FAIR Connect

Schultes, Erik. 'The FAIR Hourglass: A Framework for FAIR Implementation'. 1 Jan. 2023 : 13 – 17.
https://content.iospress.com/articles/fair-connect/fc221514

FAIR Hourglass

Raw Data
[freedom to operate]

FAIRification

Data that are FAIR-Ready
[a minimal standard that is FAIR, open, and technology-independet]

FAIR Orchestration

Data Integration
Data Sharing/Visiting
Data Analytics
[freedom to operate]

Creation of data & capture of data-reuse consent choices made by appropriate stakeholders

ETL process for data & consent targeting domain-relevant machine-readable vocabularies & schema

Storage of data (distributed) and consent as machine-actionable units of information

Exposure of FAIR metadata & gateways for ontology-based access control

Automated FAIR services for indexing, search, semantic resolution, data integration, object routing & workflow execution

High-level data use applications such as data landscape surveys, data visiting and data analytics

Wizard

CEDAR Workbench

nanobench

BioPortal

FAIR DATA POINT

R

jupyter

FAIR Connect

Schultes, Erik. 'The FAIR Hourglass: A Framework for FAIR Implementation'. 1 Jan. 2023 : 13 – 17.
https://content.iospress.com/articles/fair-connect/fc221514

FAIR Hourglass

Raw Data
[freedom to operate]

FAIRification

Data that are FAIR-Ready
[a minimal standard that is FAIR, open, and technology-independet]

FAIR Orchestration

Data Integration
Data Sharing/Visiting
Data Analytics
[freedom to operate]

Creation of data & capture of data-reuse consent choices made by appropriate stakeholders

ETL process for data & consent targeting domain-relevant machine-readable vocabularies & schema

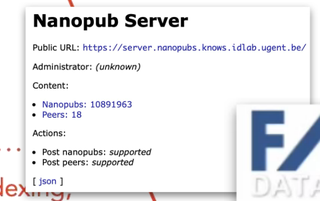
Storage of data (distributed) and consent as machine-actionable units of information

Exposure of FAIR metadata & gateways for ontology-based access control

Automated FAIR services for indexing, search, semantic resolution, data integration, object routing & workflow execution

High-level data use applications such as data landscape surveys, data visiting and data analytics

Wizard
CEDAR Workbench
nanobench
BioPortal



FAIR DATA POINT
PERSONAL HEALTH TRAIN
IOTICS
R
jupyter
FAIR Connect

Schultes, Erik. 'The FAIR Hourglass: A Framework for FAIR Implementation'. 1 Jan. 2023 : 13 – 17.
https://content.iospress.com/articles/fair-connect/fc221514

FAIR Hourglass

Raw Data
[freedom to operate]

FAIRification

Data that are FAIR-Ready
[a minimal standard that is FAIR, open, and technology-independet]

FAIR Orchestration

Data Integration
Data Sharing/Visiting
Data Analytics
[freedom to operate]

Creation of data & capture of data-reuse consent choices made by approprite stakeholders

ETL process for data & consent targeting domain-relevant machine-readable vocabularies & schema

Storage of data (distributed) and consent as machine-actionable units of information

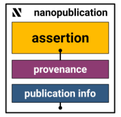
Exposure of FAIR metadata & gateways for ontology-based access control

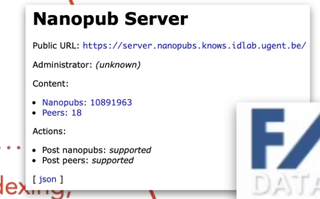
Automated FAIR services for indexing, search, semantic resolution, data integration, object routing & workflow execution

High-level data use applications such as data landscape surveys, data visiting and data analytics

Wizard CEDAR Workbench

nanobench BioPortal





FAIR DATA POINT

PERSONAL HEALTH TRAIN IOTICS

R jupyter FAIR Connect

Schultes, Erik. 'The FAIR Hourglass: A Framework for FAIR Implementation'. 1 Jan. 2023 : 13 – 17.
https://content.iospress.com/articles/fair-connect/fc221514

https://doi.org/10.3390/info12110472

01010
01010
01010 *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

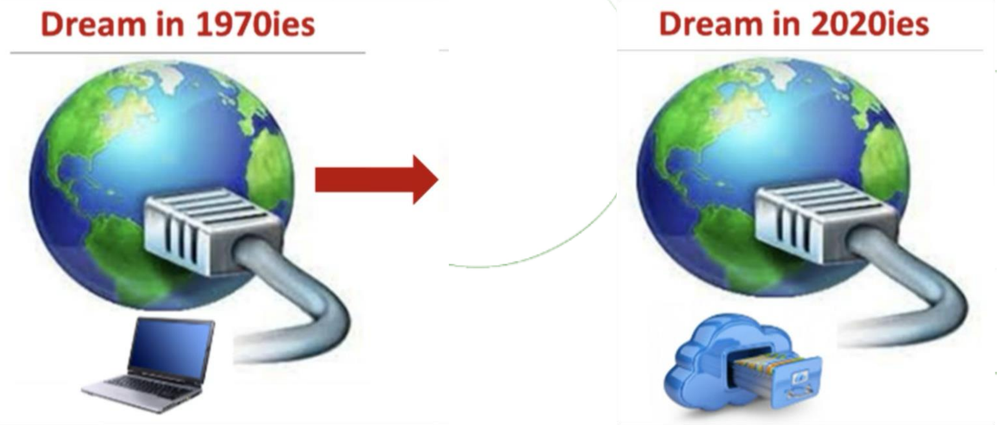
Citation: Wittenburg, P.; Strawn, G. Revolutions Take Time. *Information* **2021**, *12*, 472. https://doi.org/10.3390/info12110472

Academic Editor: Riccardo Albertoni

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

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



An internet of **computers**
A web of **documents**

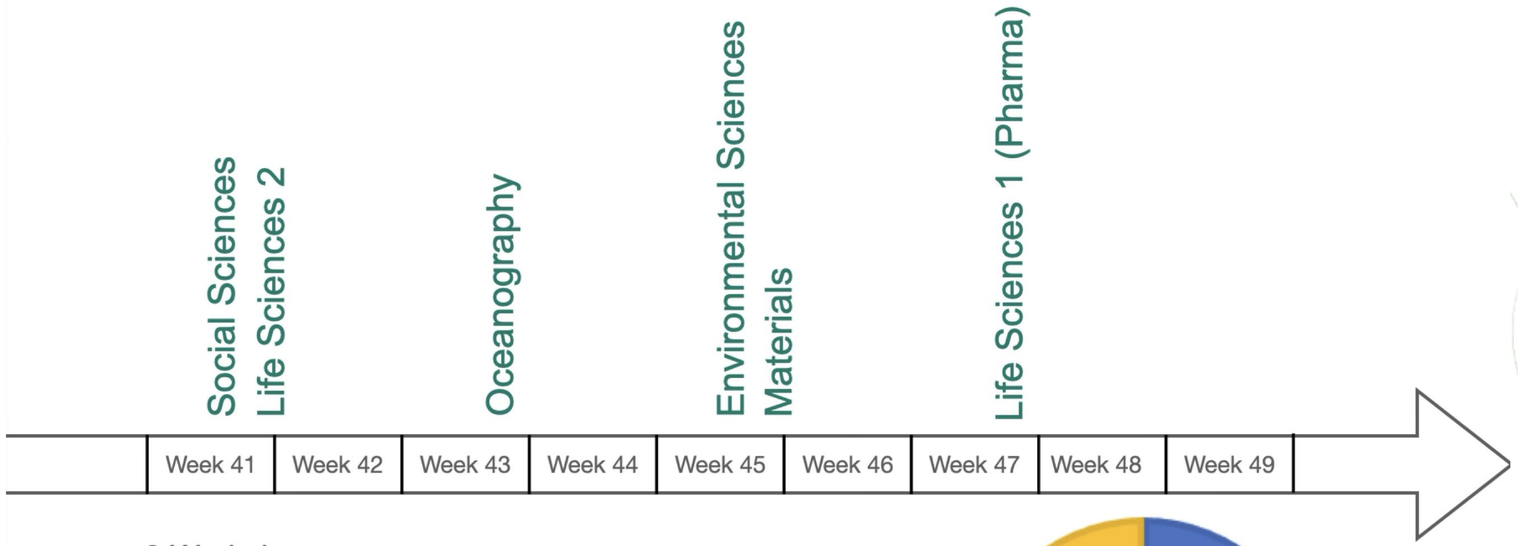
An internet of **data & services**
A web of **data**

FIP analysis:
The FIP matrix &
The Convergence matrix

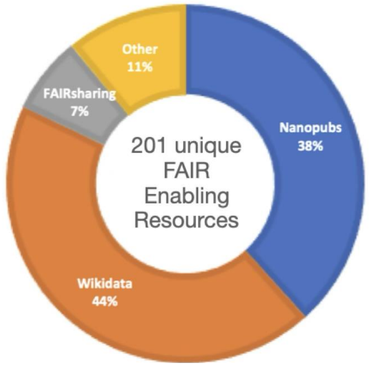
13:50-14:00
(Erik)

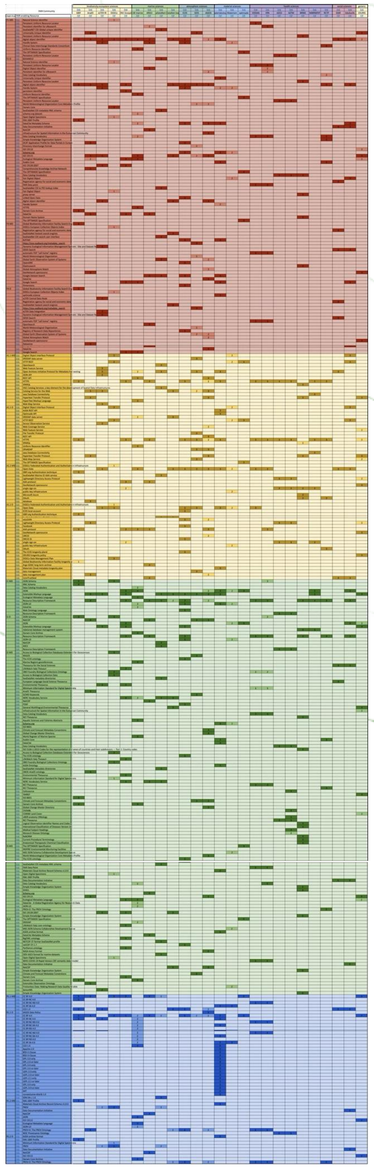
FIP matrix

2020 CODATA & GO FAIR Convergence Symposium



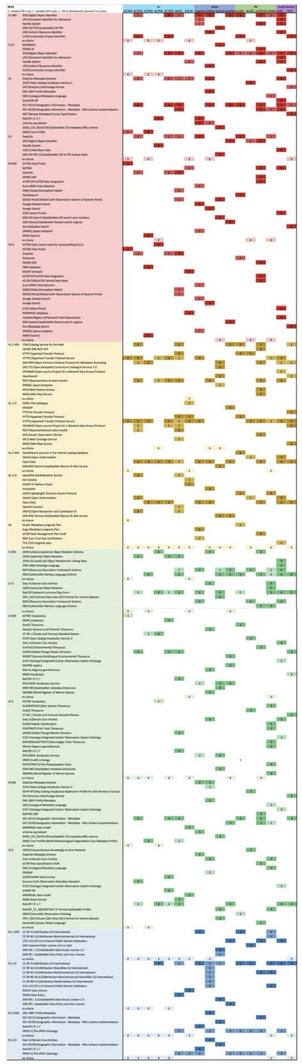
- 6 Workshops
- 25 communites
- 81 participants
- 46 contact hours
- Outputs:
  - Human-readable FIPs https://osf.io/r2hzc/
  - FIPs as graphs https://osf.io/6sbfy/
  - Convergence Matrix https://bit.ly/3lKCGFl



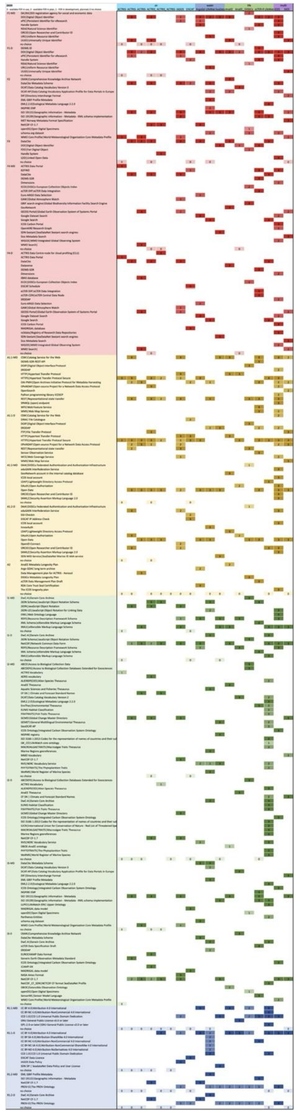


FIP matrix

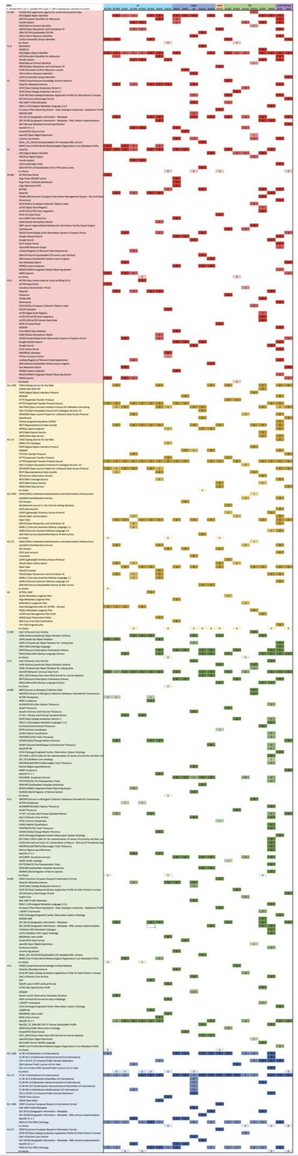
2019



2020



2021

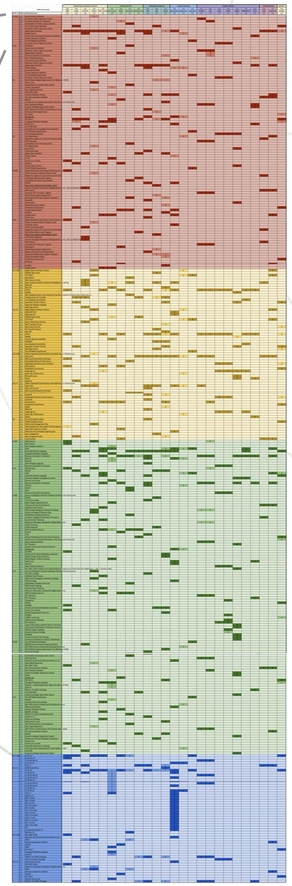


Number of RIs: 14
Number of FICs: 22
Number of assesment cycles: 3 (2019, 2020, 2021)
Total number of FIPs: 57

Total number of FERs declared: 193
Total number of triples captured: 1852
most FERs: I2 MD and D
most no choices: A2 (followed by R1.1 MD, R1.2 MD, I1 D)

FIP matrix

FAIR Prin	FAIR-enabling Resource	
F1-MD	link	Natural Science Identifier
	link	Persistent Uniform Resource Locator
	link	Persistent Identifier for eResearch
	link	SeaDataNET CDI Global unique identifier
	link	Universally Unique Identifier
	link	Persistent Uniform Resource Locator
	link	digital object identifier
	link	Handle System
	link	Clinical Data Interchange Standards Consortium
	link	Uniform Resource Identifier
	link	The OPTIMADE Specification
	link	Persistent Uniform Resource Locator
F1-D	link	B2HANDLE
	link	Natural Science Identifier
	link	Persistent Uniform Resource Locator
	link	Digital Object Identifier
	link	Persistent Identifier for eResearch
	link	Data Catalog Vocabulary
	link	Universally Unique Identifier
	link	Persistent Uniform Resource Locator
	link	digital object identifier
	link	Handle System
	link	persistent identifier
	link	Uniform Resource Identifier
	link	The OPTIMADE Specification
	link	Persistent Uniform Resource Locator
F2	link	World Meteorological Organization Core Metadata
	link	Darwin Core
	link	SeaDataNet CDI metadata XML schema
	link	schema.org dataset
	link	Open Digital Specimens
	link	EML GBIF Profile
	link	DataCite Metadata Scheme
	link	Data Documentation Initiative
	link	NetCDF
	link	Infrastructure for Spatial Information in the Europ
	link	Data Catalog Vocabulary
	link	Simple Knowledge Organization System
	link	DCAT Application Profile for Data Portals in Europe
	link	Directory Interchange Format
	link	ISO 19110
	link	Schema.org
	link	ISO 19115
	link	Ecological Metadata Language
	link	Dublin Core
	link	ISO 19139:2007
	link	Comprehensive Knowledge Archive Network
	link	The OPTIMADE Specification



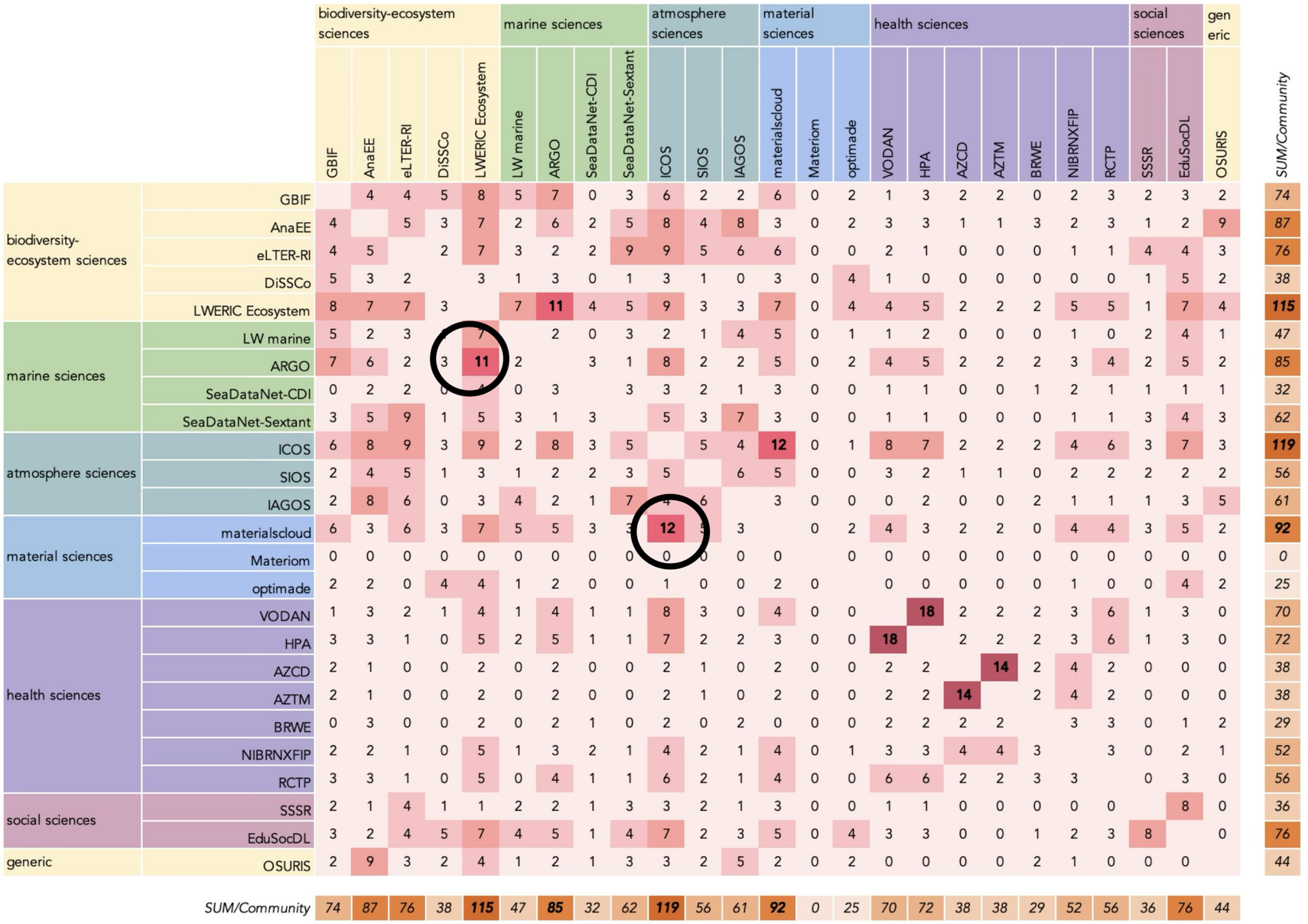
Convergence matrix (FIP overlap)

		GBIF	AnaEE	eLTER-RI	DiSSCo	LWERIC Ecosystem	LW marine	ARGO	SeaDataNet-CDI	SeaDataNet-Sextant	ICOS	SIOS	IAGOS	materialscloud	Materiom	optimade	VODAN	HPA	AZCD	AZTM	BRWE	NIBRNXFIP	RCTP	SSSR	EduSocDL	OSURIS	*SUM/Community*
		biodiversity-ecosystem sciences					marine sciences				atmosphere sciences			material sciences			health sciences							social sciences		generic	
biodiversity-ecosystem sciences	GBIF		4	4	5	8	5	7	0	3	6	2	2	6	0	2	1	3	2	2	0	2	3	2	3	2	*74*
	AnaEE	4		5	3	7	2	6	2	5	8	4	8	3	0	2	3	3	1	1	3	2	3	1	2	9	*87*
	eLTER-RI	4	5		2	7	3	2	2	9	9	5	6	6	0	0	2	1	0	0	0	1	1	4	4	3	*76*
	DiSSCo	5	3	2		3	1	3	0	1	3	1	0	3	0	4	1	0	0	0	0	0	0	1	5	2	*38*
	LWERIC Ecosystem	8	7	7	3		7	**11**	4	5	9	3	3	7	0	4	4	5	2	2	2	5	5	1	7	4	***115***
marine sciences	LW marine	5	2	3	1	7		2	0	3	2	1	4	5	0	1	1	2	0	0	0	1	0	2	4	1	*47*
	ARGO	7	6	2	3	**11**	2		3	1	8	2	2	5	0	2	4	5	2	2	2	3	4	2	5	2	*85*
	SeaDataNet-CDI	0	2	2	0	4	0	3		3	3	2	1	3	0	0	1	1	0	0	1	2	1	1	1	1	*32*
	SeaDataNet-Sextant	3	5	9	1	5	3	1	3		5	3	7	3	0	0	1	1	0	0	0	1	1	3	4	3	*62*
atmosphere sciences	ICOS	6	8	9	3	9	2	8	3	5		5	4	**12**	0	1	8	7	2	2	2	4	6	3	7	3	***119***
	SIOS	2	4	5	1	3	1	2	2	3	5		6	5	0	0	3	2	1	1	0	2	2	2	2	2	*56*
	IAGOS	2	8	6	0	3	4	2	1	7	4	6		3	0	0	0	2	0	0	2	1	1	1	3	5	*61*
material sciences	materialscloud	6	3	6	3	7	5	5	3	3	**12**	5	3		0	2	4	3	2	2	0	4	4	3	5	2	***92***
	Materiom	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	*0*
	optimade	2	2	0	4	4	1	2	0	0	1	0	0	2	0		0	0	0	0	0	1	0	0	4	2	*25*
health sciences	VODAN	1	3	2	1	4	1	4	1	1	8	3	0	4	0	0		**18**	2	2	2	3	6	1	3	0	*70*
	HPA	3	3	1	0	5	2	5	1	1	7	2	2	3	0	0	**18**		2	2	2	3	6	1	3	0	*72*
	AZCD	2	1	0	0	2	0	2	0	0	2	1	0	2	0	0	2	2		**14**	2	4	2	0	0	0	*38*
	AZTM	2	1	0	0	2	0	2	0	0	2	1	0	2	0	0	2	2	**14**		2	4	2	0	0	0	*38*
	BRWE	0	3	0	0	2	0	2	1	0	2	0	2	0	0	0	2	2	2	2		3	3	0	1	2	*29*
	NIBRNXFIP	2	2	1	0	5	1	3	2	1	4	2	1	4	0	1	3	3	4	4	3		3	0	2	1	*52*
	RCTP	3	3	1	0	5	0	4	1	1	6	2	1	4	0	0	6	6	2	2	3	3		0	3	0	*56*
social sciences	SSSR	2	1	4	1	1	2	2	1	3	3	2	1	3	0	0	1	1	0	0	0	0	0		8	0	*36*
	EduSocDL	3	2	4	5	7	4	5	1	4	7	2	3	5	0	4	3	3	0	0	1	2	3	8		0	*76*
generic	OSURIS	2	9	3	2	4	1	2	1	3	3	2	5	2	0	2	0	0	0	0	2	1	0	0	0		*44*
	SUM/Community	*74*	***87***	*76*	*38*	***115***	*47*	***85***	*32*	*62*	***119***	*56*	*61*	***92***	*0*	*25*	*70*	*72*	*38*	*38*	*29*	*52*	*56*	*36*	***76***	*44*	

Convergence matrix (FIP overlap)

		GBIF	AnaEE	eLTER-RI	DiSSCo	LWERIC Ecosystem	LW marine	ARGO	SeaDataNet-CDI	SeaDataNet-Sextant	ICOS	SIOS	IAGOS	materialscloud	Materiom	optimade	VODAN	HPA	AZCD	AZTM	BRWE	NIBRNXFIP	RCTP	SSSR	EduSocDL	OSURIS	SUM/Community
		biodiversity-ecosystem sciences					marine sciences				atmosphere sciences			material sciences			health sciences							social sciences		generic	
biodiversity-ecosystem sciences	GBIF		4	4	5	8	5	7	0	3	6	2	2	6	0	2	1	3	2	2	0	2	3	2	3	2	74
	AnaEE	4		5	3	7	2	6	2	5	8	4	8	3	0	2	3	3	1	1	3	2	3	1	2	9	87
	eLTER-RI	4	5		2	7	3	2	2	9	9	5	6	6	0	0	2	1	0	0	0	1	1	4	4	3	76
	DiSSCo	5	3	2		3	1	3	0	1	3	1	0	3	0	4	1	0	0	0	0	0	0	1	5	2	38
	LWERIC Ecosystem	8	7	7	3		7	**11**	4	5	9	3	3	7	0	4	4	5	2	2	2	5	5	1	7	4	**115**
marine sciences	LW marine	5	2	3	1	7		2	0	3	2	1	4	5	0	1	1	2	0	0	0	1	0	2	4	1	47
	ARGO	7	6	2	3	**11**	2		3	1	8	2	2	5	0	2	4	5	2	2	2	3	4	2	5	2	85
	SeaDataNet-CDI	0	2	2	0	4	0	3		3	3	2	1	3	0	0	1	1	0	0	1	2	1	1	1	1	32
	SeaDataNet-Sextant	3	5	9	1	5	3	1	3		5	3	7	3	0	0	1	1	0	0	0	1	1	3	4	3	62
atmosphere sciences	ICOS	6	8	9	3	9	2	8	3	5		5	4	**12**	0	1	8	7	2	2	2	4	6	3	7	3	**119**
	SIOS	2	4	5	1	3	1	2	2	3	5		6	5	0	0	3	2	1	1	0	2	2	2	2	2	56
	IAGOS	2	8	6	0	3	4	2	1	7	4	6		3	0	0	0	2	0	0	2	1	1	1	3	5	61
material sciences	materialscloud	6	3	6	3	7	5	5	3	3	**12**	5	3		0	2	4	3	2	2	0	4	4	3	5	2	**92**
	Materiom	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	*0*
	optimade	2	2	0	4	4	1	2	0	0	1	0	0	2	0		0	0	0	0	0	1	0	0	4	2	25
health sciences	VODAN	1	3	2	1	4	1	4	1	1	8	3	0	4	0	0		**18**	2	2	2	3	6	1	3	0	70
	HPA	3	3	1	0	5	2	5	1	1	7	2	2	3	0	0	**18**		2	2	2	3	6	1	3	0	72
	AZCD	2	1	0	0	2	0	2	0	0	2	1	0	2	0	0	2	2		**14**	2	4	2	0	0	0	38
	AZTM	2	1	0	0	2	0	2	0	0	2	1	0	2	0	0	2	2	**14**		2	4	2	0	0	0	38
	BRWE	0	3	0	0	2	0	2	1	0	2	0	2	0	0	0	2	2	2	2		3	3	0	1	2	29
	NIBRNXFIP	2	2	1	0	5	1	3	2	1	4	2	1	4	0	1	3	3	4	4	3		3	0	2	1	52
	RCTP	3	3	1	0	5	0	4	1	1	6	2	1	4	0	0	6	6	2	2	3	3		0	3	0	56
social sciences	SSSR	2	1	4	1	1	2	2	1	3	3	2	1	3	0	0	1	1	0	0	0	0	0		8	0	36
	EduSocDL	3	2	4	5	7	4	5	1	4	7	2	3	5	0	4	3	3	0	0	1	2	3	8		0	76
generic	OSURIS	2	9	3	2	4	1	2	1	3	3	2	5	2	0	2	0	0	0	0	2	1	0	0	0		44
	SUM/Community	*74*	*87*	*76*	*38*	***115***	*47*	***85***	*32*	*62*	***119***	*56*	*61*	***92***	*0*	*25*	*70*	*72*	*38*	*38*	*29*	*52*	*56*	*36*	*76*	*44*	

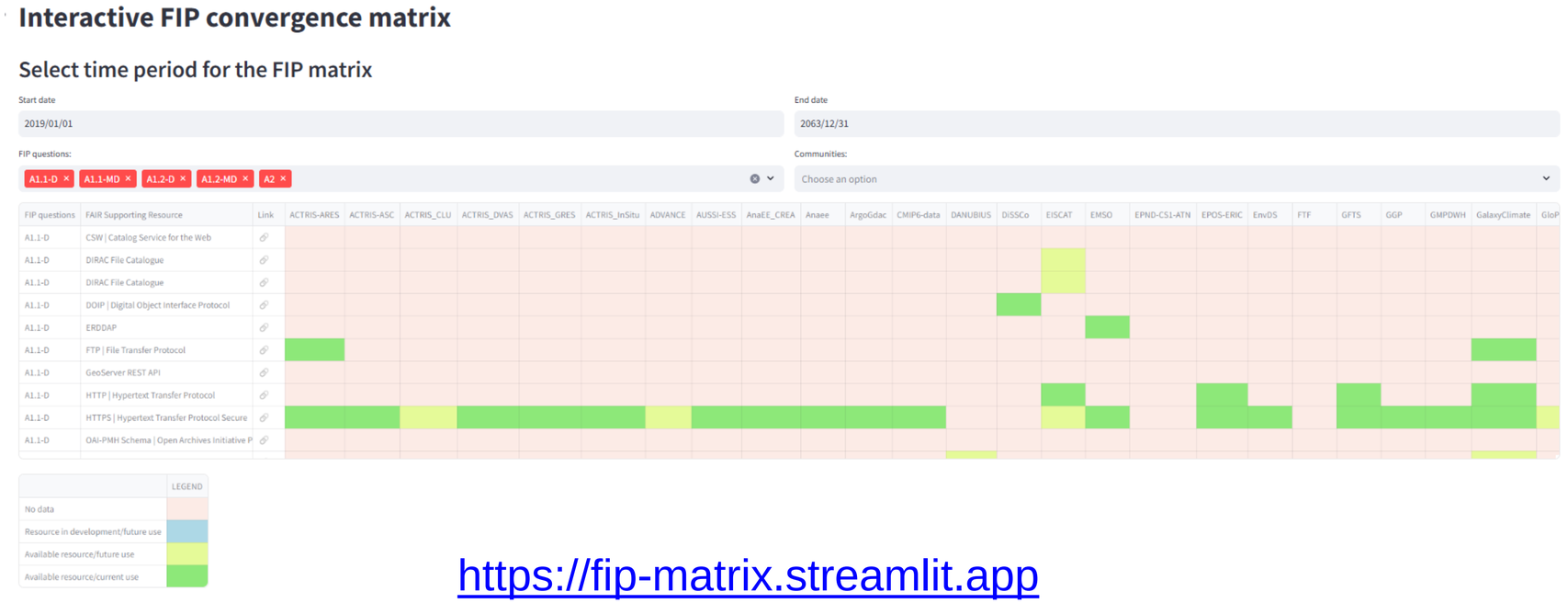
Convergence matrix (FIP overlap)



		GBIF	AnaEE	eLTER-RI	DiSSCo	LWERIC Ecosystem	LW marine	ARGO	SeaDataNet-CDI	SeaDataNet-Sextant	ICOS	SIOS	IAGOS	materialscloud	Materiom	optimade	VODAN	HPA	AZCD	AZTM	BRWE	NIBRNXFIP	RCTP	SSSR	EduSocDL	OSURIS	SUM/Community
biodiversity-ecosystem sciences	GBIF		4	4	5	8	5	7	0	3	6	2	2	6	0	2	1	3	2	2	0	2	3	2	3	2	74
	AnaEE	4		5	3	7	2	6	2	5	8	4	8	3	0	2	3	3	1	1	3	2	3	1	2	9	87
	eLTER-RI	4	5		2	7	3	2	2	9	9	5	6	6	0	0	2	1	0	0	0	1	1	4	4	3	76
	DiSSCo	5	3	2		3	1	3	0	1	3	1	0	3	0	4	1	0	0	0	0	0	0	1	5	2	38
	LWERIC Ecosystem	8	7	7	3		7	11	4	5	9	3	3	7	0	4	4	5	2	2	2	5	5	1	7	4	115
marine sciences	LW marine	5	2	3	1	7		2	0	3	2	1	4	5	0	1	1	2	0	0	0	1	0	2	4	1	47
	ARGO	7	6	2	3	11	2		3	1	8	2	2	5	0	2	4	5	2	2	2	3	4	2	5	2	85
	SeaDataNet-CDI	0	2	2	0	1	0	3		3	3	2	1	3	0	0	1	1	0	0	1	2	1	1	1	1	32
	SeaDataNet-Sextant	3	5	9	1	5	3	1	3		5	3	7	3	0	0	1	1	0	0	0	1	1	3	4	3	62
atmosphere sciences	ICOS	6	8	9	3	9	2	8	3	5		5	4	12	0	1	8	7	2	2	2	4	6	3	7	3	119
	SIOS	2	4	5	1	3	1	2	2	3	5		6	5	0	0	3	2	1	1	0	2	2	2	2	2	56
	IAGOS	2	8	6	0	3	4	2	1	7	4	6		3	0	0	0	2	0	0	2	1	1	1	3	5	61
material sciences	materialscloud	6	3	6	3	7	5	5	3	3	12	3	3		0	2	4	3	2	2	0	4	4	3	5	2	92
	Materiom	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0
	optimade	2	2	0	4	4	1	2	0	0	1	0	0	2	0		0	0	0	0	0	1	0	0	4	2	25
health sciences	VODAN	1	3	2	1	4	1	4	1	1	8	3	0	4	0	0		18	2	2	2	3	6	1	3	0	70
	HPA	3	3	1	0	5	2	5	1	1	7	2	2	3	0	0	18		2	2	2	3	6	1	3	0	72
	AZCD	2	1	0	0	2	0	2	0	0	2	1	0	2	0	0	2	2		14	2	4	2	0	0	0	38
	AZTM	2	1	0	0	2	0	2	0	0	2	1	0	2	0	0	2	2	14		2	4	2	0	0	0	38
	BRWE	0	3	0	0	2	0	2	1	0	2	0	2	0	0	0	2	2	2	2		3	3	0	1	2	29
	NIBRNXFIP	2	2	1	0	5	1	3	2	1	4	2	1	4	0	1	3	3	4	4	3		3	0	2	1	52
	RCTP	3	3	1	0	5	0	4	1	1	6	2	1	4	0	0	6	6	2	2	3	3		0	3	0	56
social sciences	SSSR	2	1	4	1	1	2	2	1	3	3	2	1	3	0	0	1	1	0	0	0	0	0		8	0	36
	EduSocDL	3	2	4	5	7	4	5	1	4	7	2	3	5	0	4	3	3	0	0	1	2	3	8		0	76
generic	OSURIS	2	9	3	2	4	1	2	1	3	3	2	5	2	0	2	0	0	0	0	2	1	0	0	0		44
	SUM/Community	74	87	76	38	115	47	85	32	62	119	56	61	92	0	25	70	72	38	38	29	52	56	36	76	44	

Breakout: Explore the FIP matrix

14:00-14:20



https://fip-matrix.streamlit.app

FIP analysis: statistical features

14:35-14:55
(Erik)

D5.6
Synthesis and future strategy

Work Package	WP5
Lead partner	EAA
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	30 June 2023
Submission date	30 June 2023

Deliverable abstract

ENVRI-FAIR is working towards the integration of (meta)data and service flows across the different environmental research infrastructures and to enhance the FAIRness of their (meta)data and services and enable efficient linking to EOSC. In this respect, FAIR Convergence of the participating research infrastructures will enable a tight integration of services and datasets from the participating ENVRIs, to the benefit of various stakeholder categories. Convergence, both technological and conceptual, will support integration with the European Open Science Cloud. The current report D5.6 on *Synthesis and outlook* evaluates the current status of the FAIR convergence efforts done on the subdomain and cluster level and provides strategic recommendations for the future developments of the ENVRIs and the ENVRI-Hub.

Peterseil, J., Offenthaler, I., Wohner, C., Magagna, B., Schultes, E., Lund Myhre, C., Jeffery, K., Bailo, D., Dobler, D., Portier, M., Dema, C., Vaira, L., & Rosati, I. (2023). ENVRI-FAIR D5.6: Synthesis and future strategy (Version 1). Zenodo. https://doi.org/10.5281/zenodo.8118948

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

ITINERIS

Project Title	Global cooperation on FAIR data policy and practice
Project Acronym	WorldFAIR
Grant Agreement No	101058393
Instrument	HORIZON-WIDERA-2021-ERA-01
Topic, type of action	HORIZON-WIDERA-2021-ERA-01-41 HORIZON Coordination and Support Actions
Start Date of Project	2022-06-01
Duration of Project	24 months
Project Website	http://worldfair-project.eu

D2.1 'FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?'

Deliverable	'FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?' report on D2.1 'Completed FAIR Implementation Profiles for each Case Study'
Work Package	WP02 - Recommendations, Synthesis, FAIR Assessment
Lead Author (Org)	Arofan Gregory (CODATA) and Simon Hodson (CODATA)
Contributing Author(s) (Org)	Lesley Wyborn (Australian National University) Laura Molloy (CODATA)

Gregory, A., & Hodson, S. (2022). WorldFAIR Project (D2.1) 'FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?' (1.0). Zenodo. https://doi.org/10.5281/zenodo.7378109

ENVRI-FAIR DELIVERABLE

D5.6
Synthesis and future strategy

Work Package	WP5
Lead partner	EAA
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	30 June 2023
Submission date	30 June 2023

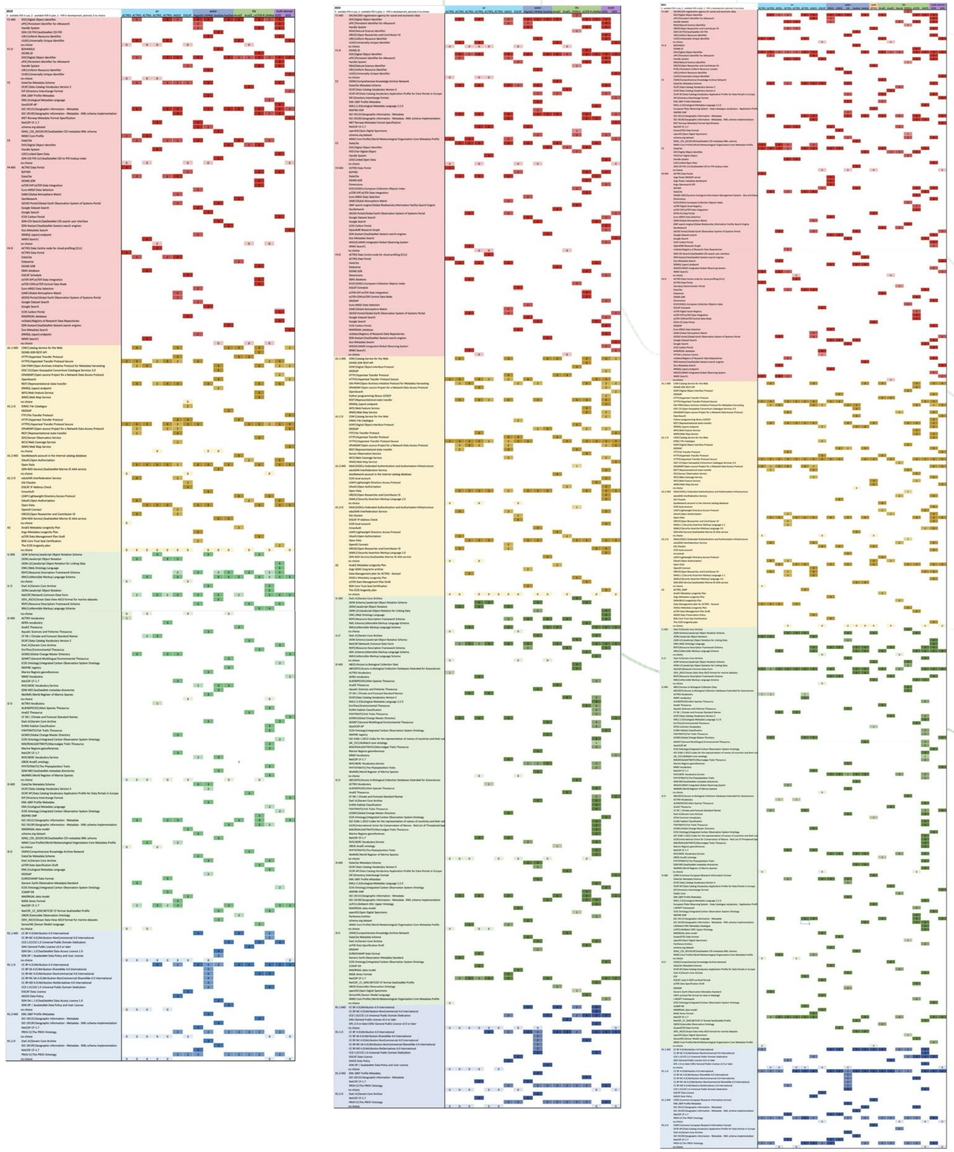
Deliverable abstract

ENVRI-FAIR is working towards the integration of (meta)data and service flows across the different environmental research infrastructures and to enhance the FAIRness of their (meta)data and services and enable efficient linking to EOSC. In this respect, FAIR Convergence of the participating research infrastructures will enable a tight integration of services and datasets from the participating ENVRIs, to the benefit of various stakeholder categories. Convergence, both technological and conceptual, will support integration with the European Open Science Cloud. The current report D5.6 on *Synthesis and outlook* evaluates the current status of the FAIR convergence efforts done on the subdomain and cluster level and provides strategic recommendations for the future developments of the ENVRIs and the ENVRI-Hub.

Peterseil, J., Offenthaler, I., Wohner, C., Magagna, B., Schultes, E., Lund Myhre, C., Jeffery, K., Bailo, D., Dobler, D., Portier, M., Dema, C., Vaira, L., & Rosati, I. (2023). ENVRI-FAIR D5.6: Synthesis and future strategy (Version 1). Zenodo. https://doi.org/10.5281/zenodo.8118948

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

2019 2020 2021



ITINERIS

D5.6
Synthesis and future strategy

Work Package	WP5
Lead partner	EAA
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	30 June 2023
Submission date	30 June 2023

Deliverable abstract

ENVRI-FAIR is working towards the integration of (meta)data and service flows across the different environmental research infrastructures and to enhance the FAIRness of their (meta)data and services and enable efficient linking to EOSC. In this respect, FAIR Convergence of the participating research infrastructures will enable a tight integration of services and datasets from the participating ENVRIs, to the benefit of various stakeholder categories. Convergence, both technological and conceptual, will support integration with the European Open Science Cloud. The current report D5.6 on *Synthesis and outlook* evaluates the current status of the FAIR convergence efforts done on the subdomain and cluster level and provides strategic recommendations for the future developments of the ENVRIs and the ENVRI-Hub.

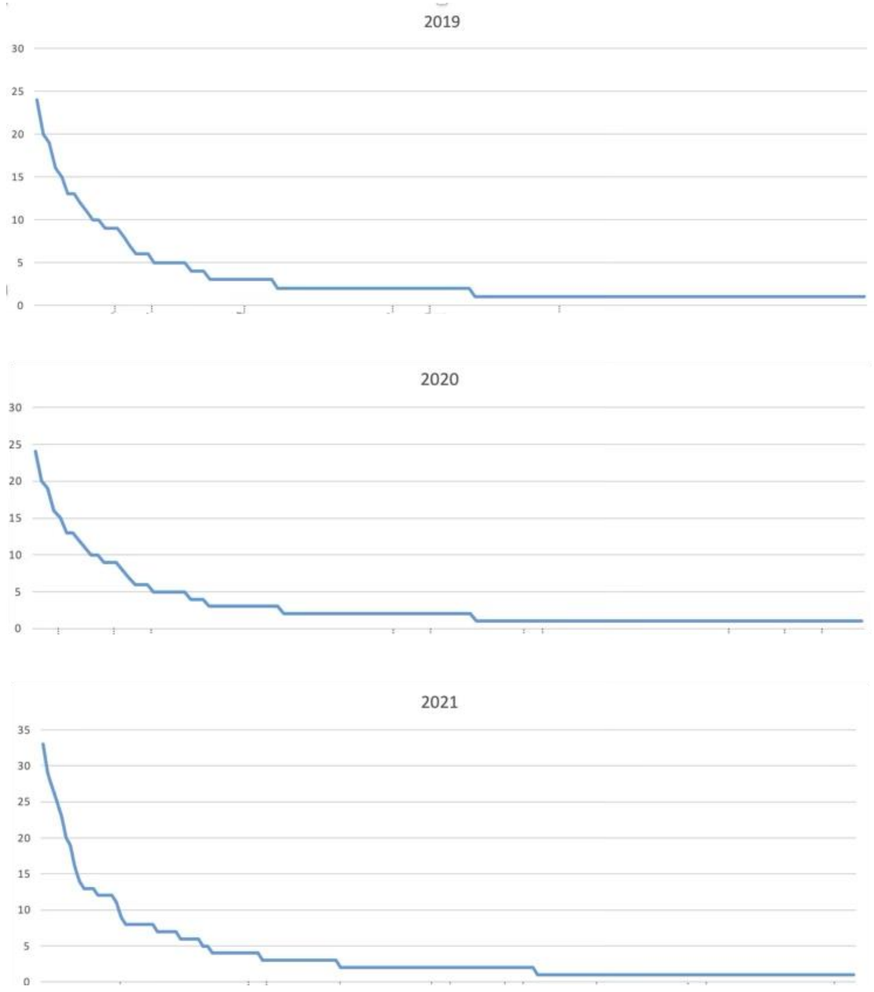
Peterseil, J., Offenthaler, I., Wohner, C., Magagna, B., Schultes, E., Lund Myhre, C., Jeffery, K., Bailo, D., Dobler, D., Portier, M., Dema, C., Vaira, L., & Rosati, I. (2023). ENVRI-FAIR D5.6: Synthesis and future strategy (Version 1). Zenodo. https://doi.org/10.5281/zenodo.8118948

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

ITINERIS

Number of RIs: 14
Number of FICs: 22
Number of assesment cycles: 3 (2019, 2020, 2021)
Total number of FIPs: 57

Total number of FERs declared: 193
Total number of triples captured: 1852
most FERs: I2 MD and D
most no choices: A2 (followed by R1.1 MD, R1.2 MD, I1 D)



DOI|Digital Object Identifier	33
CC BY 4.0|Attribution 4.0 International	29
HTTPS|Hypertext Transfer Protocol Secure	27
PROV-O|The PROV Ontology	25
NetCDF CF-1.7	23
DataCite	20
Open Data	19
ISO 19115|Geographic information - Metadata	16
NetCDF|Network Common Data Form	14
Handle System	13
NVS|NERC Vocabulary Service	13
REST|Representational state transfer	13
DataCite Metadata Scheme	12
ORCID|Open Researcher and Contributor ID	12
RDFS|Resource Description Framework Schema	12
XMLS|eXtensible Markup Language Schema	12
OPeNDAP|Open-source Project for a Network Data Access Protocol	11

Number FERs per FAIR Principle

FAIR principle	Question	FER type	FER count
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records/datasets?	Identifier type	10
F2	Which metadata schemas do you use for findability?	Metadata schema	19
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	6
F4	In which search engines are your metadata records/datasets indexed?	Search engines	28
A1.1	Which standardized communication protocol do you use for metadata records/datasets?	Communication protocol	15
A1.2	Which authentication & authorisation technique do you use for metadata records/datasets?	Authentication & authorisation technique	13
A2	Which metadata longevity plan do you use?	Metadata longevity	10
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records/datasets?	Knowledge representation language	7
I2	Which structured vocabularies do you use to annotate your metadata records/datasets?	Structured vocabularies	29
I3	Which models, schema(s) do you use for your datasets?	Data schema	23
R1.1	Which usage license do you use for your metadata records/datasets?	Data usage license	8
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records/datasets?	Provenance model	6

ENVRI FAIR

Total FERs = 174

Average FERs per Principle = 15

Number FERs per FAIR Principle

FAIR principle	Question	FER type	FER count
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records/datasets?	Identifier type	10
F2	Which metadata schemas do you use for findability?	Metadata schema	19
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	6
F4	In which search engines are your metadata records/datasets indexed?	Search engines	28
A1.1	Which standardized communication protocol do you use for metadata records/datasets?	Communication protocol	15
A1.2	Which authentication & authorisation technique do you use for metadata records/datasets?	Authentication & authorisation technique	13
A2	Which metadata longevity plan do you use?	Metadata longevity	10
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records/datasets?	Knowledge representation language	7
I2	Which structured vocabularies do you use to annotate your metadata records/datasets?	Structured vocabularies	29
I3	Which models, schema(s) do you use for your datasets?	Data schema	23
R1.1	Which usage license do you use for your metadata records/datasets?	Data usage license	8
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records/datasets?	Provenance model	6

Total FERs = 288

Average FERs per Principle = 19

Number FERs per community (FIP length)

	ACTRIS-DVAS	ACTRIS-GRES	ACTRIS-InSitu	ACTRIS_ARES	ACTRIS_CLU	ACTRIS-ASC	IAGOS	EISCAT	ArgoGdaC	EMSO ERIC	LifeWatch_Marine	SeaDataNET_CDI	SeaDataNET_Sextant	EPOS_ICS-C	AnaEE	AnaEE_CREA	DANUBIUS	DiSSCo	eLTER-RI	LifeWatch_Ecosystems	ICOS	SIOS	
2019	6	25	11		13	22	41	14	32		33	29	21		23	17			39	31	43	26	25,1
2020	8	28	21	29	14		47	17	30		41	23			27	17		31	42	56	68	32	31,2
2021	19	32	24	38	20	28	46	19	35	22	39	30	29	23	27	17	22	30	44	56	66	30	31,6
	11,0	28,3	18,7	33,5	15,7	16,7	44,7	16,7	32,3	22,0	37,7	27,3	25,0	11,5	25,7	17,0	22,0	30,5	41,7	47,7	59,0	29,3	

Maximum = 68

Average = 28

FIP Overlap 2019

	ACTRIS_CLU	ACTRIS_DVAS	ACTRIS_GRES	ACTRIS_InSitu	ACTRIS_ASC	IAGOS	EISCAT	ArgoGdac2	lw-marine	SeaDataNet-CDI	SeaDataNet-Sextant	AnaEE_CREA	Anaee	LWERIC_Ecosystem	eLTER-RI	ICOS	SIOS	
ACTRIS_CLU		2	5	3	5	6	2	6	4	2	2	2	2	3	5	7	6	62
ACTRIS_DVAS	2		1	2	1	2	1	1				1		1	1	1	1	15
ACTRIS_GRES	5	1		3	13	15	2	7	5	2	5	3	6	5	7	11	11	101
ACTRIS_InSitu	3	2	3		2	4	1	2	1	2	2	1	1	2	4	4	5	39
ACTRIS_ASC	5	1	13	2		12	2	7	5	1	4	2	5	5	5	9	9	87
IAGOS	6	2	15	4	12		4	10	7	6	9	6	9	9	9	14	11	133
EISCAT	2	1	2	1	2	4		4	2		1	3	2	1	4	5	1	35
ArgoGdac	6	1	7	2	7	10	4		6	5	6	5	5	6	5	11	6	92
lw-marine	4		5	1	5	7	2	6		1	6	4	4	7	8	10	4	74
SeaDataNet-CDI	2		2	2	1	6		5	1		6	2	3	6	3	6	5	50
SeaDataNet-Sextant	2		5	2	4	9	1	6	6	6		3	5	5	8	9	6	77
AnaEE_CREA	2	1	3	1	2	6	3	5	4	2	3		5	4	3	6	3	53
Anaee	2		6	1	5	9	2	5	4	3	5	5		6	5	6	7	71
LWERIC_Ecosystem	3	1	5	2	5	9	1	6	7	6	5	4	6		6	9	5	80
eLTER-RI	5	1	7	4	5	9	4	5	8	3	8	3	5	6		12	8	93
ICOS	7	1	11	4	9	14	5	11	10	6	9	6	6	9	12		10	130
SIOS	6	1	11	5	9	11	1	6	4	5	6	3	7	5	8	10		98
	62	15	101	39	87	133	35	92	74	50	77	53	71	80	93	130	98	1290

ACTRIS_DVAS	2		1	2	1	2	1	1				1		1	1	1	1	15
EISCAT	2	1	2	1	2	4		4	2		1	3	2	1	4	5	1	35
ACTRIS_InSitu	3	2	3		2	4	1	2	1	2	2	1	1	2	4	4	5	39
SeaDataNet-CDI	2		2	2	1	6		5	1		6	2	3	6	3	6	5	50
AnaEE_CREA	2	1	3	1	2	6	3	5	4	2	3		5	4	3	6	3	53
ACTRIS_CLU		2	5	3	5	6	2	6	4	2	2	2	2	3	5	7	6	62
Anaee	2		6	1	5	9	2	5	4	3	5	5		6	5	6	7	71
lw-marine	4		5	1	5	7	2	6		1	6	4	4	7	8	10	4	74
SeaDataNet-Sextant	2		5	2	4	9	1	6	6	6		3	5	5	8	9	6	77
LWERIC_Ecosystem	3	1	5	2	5	9	1	6	7	6	5	4	6		6	9	5	80
ACTRIS_ASC	5	1	13	2		12	2	7	5	1	4	2	5	5	5	9	9	87
ArgoGdac	6	1	7	2	7	10	4		6	5	6	5	5	6	5	11	6	92
eLTER-RI	5	1	7	4	5	9	4	5	8	3	8	3	5	6		12	8	93
SIOS	6	1	11	5	9	11	1	6	4	5	6	3	7	5	8	10		98
ACTRIS_GRES	5	1		3	13	15	2	7	5	2	5	3	6	5	7	11	11	101
ICOS	7	1	11	4	9	14	5	11	10	6	9	6	6	9	12		10	130
IAGOS	6	2	15	4	12		4	10	7	6	9	6	9	9	9	14	11	133

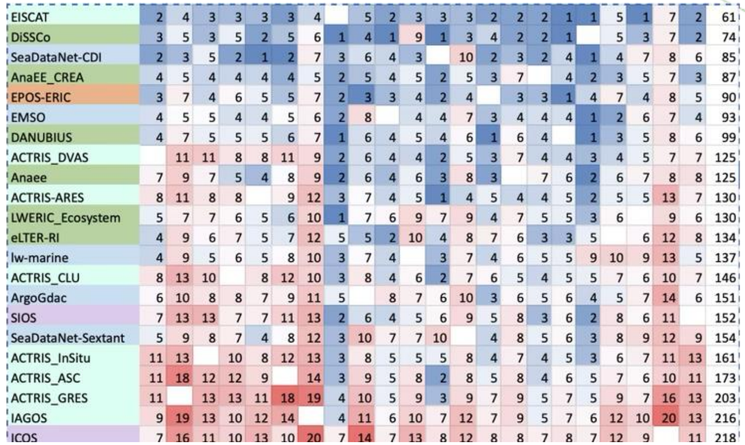
FIP Overlap 2020

	ACTRIS_DVAS	ACTRIS_GRES	ACTRIS_InSitu	ACTRIS-ARES	ACTRIS_CLU	ACTRIS_ASC	IAGOS	EISCAT	ArgoGdac2	lw-marine	SeaDataNet-Sextant	LWERIC_Ecosystem	Anaee	AnaEE_CREA	DiSSCo	eLTER-RI	ICOS	SIOS	
ACTRIS_DVAS		1	2	1	2	1	2	1	1			1	1	1		1	2	2	19
ACTRIS_GRES	1		8	6	6	13	16	3	8	7	5	5	7	3	3	8	14	11	124
ACTRIS_InSitu	2	8		4	5	7	10	2	6	3	4	6	5	2	1	5	9	12	91
ACTRIS-ARES	1	6	4		5	6	6	1	3	2	1	2	2	1	2	3	5	5	55
ACTRIS_CLU	2	6	5	5		6	7	2	6	4	2	4	3	2	4	5	7	6	76
ACTRIS_ASC	1	13	7	6	6		12	2	7	6	4	4	6	2	3	6	8	9	102
IAGOS	2	16	10	6	7	12		3	10	9	10	9	9	5	6	11	19	13	157
EISCAT	1	3	2	1	2	2	3		3	3	1	1	2	2	1	5	6	2	40
ArgoGdac	1	8	6	3	6	7	10	3		6	5	6	5	4	3	4	12	6	95
lw-marine		7	3	2	4	6	9	3	6		7	9	5	5	8	11	13	5	103
SeaDataNet-Sextant		5	4	1	2	4	10	1	5	7		6	5	3	3	9	9	6	80
LWERIC_Ecosystem	1	5	6	2	4	4	9	1	6	9	6		7	4	3	7	7	6	87
Anaee	1	7	5	2	3	6	9	2	5	5	5	7		7	2	6	8	8	88
AnaEE_CREA	1	3	2	1	2	2	5	2	4	5	3	4	7		2	3	7	3	56
DiSSCo		3	1	2	4	3	6	1	3	8	3	3	2	2		5	6	2	54
eLTER-RI	1	8	5	3	5	6	11	5	4	11	9	7	6	3	5		13	8	110
ICOS	2	14	9	5	7	8	19	6	12	13	9	7	8	7	6	13		12	157
SIOS	2	11	12	5	6	9	13	2	6	5	6	6	8	3	2	8	12		116
	19	124	91	55	76	102	157	40	95	103	80	87	88	56	54	110	157	116	1610

ACTRIS_DVAS		1	2	1	2	1	2	1	1			1	1	1		1	2	2	19
EISCAT	1	3	2	1	2	2	3		3	3	1	1	2	2	1	5	6	2	40
DiSSCo		3	1	2	4	3	6	1	3	8	3	3	2	2		5	6	2	54
ACTRIS-ARES	1	6	4		5	6	6	1	3	2	1	2	2	1	2	3	5	5	55
AnaEE_CREA	1	3	2	1	2	2	5	2	4	5	3	4	7		2	3	7	3	56
ACTRIS_CLU	2	6	5	5		6	7	2	6	4	2	4	3	2	4	5	7	6	76
SeaDataNet-Sextant		5	4	1	2	4	10	1	5	7		6	5	3	3	9	9	6	80
LWERIC_Ecosystem	1	5	6	2	4	4	9	1	6	9	6		7	4	3	7	7	6	87
Anaee	1	7	5	2	3	6	9	2	5	5	5	7		7	2	6	8	8	88
ACTRIS_InSitu	2	8		4	5	7	10	2	6	3	4	6	5	2	1	5	9	12	91
ArgoGdac	1	8	6	3	6	7	10	3		6	5	6	5	4	3	4	12	6	95
ACTRIS_ASC	1	13	7	6	6		12	2	7	6	4	4	6	2	3	6	8	9	102
lw-marine		7	3	2	4	6	9	3	6		7	9	5	5	8	11	13	5	103
eLTER-RI	1	8	5	3	5	6	11	5	4	11	9	7	6	3	5		13	8	110
SIOS	2	11	12	5	6	9	13	2	6	5	6	6	8	3	2	8	12		116
ACTRIS_GRES	1		8	6	6	13	16	3	8	7	5	5	7	3	3	8	14	11	124
IAGOS	2	16	10	6	7	12		3	10	9	10	9	9	5	6	11	19	13	157
ICOS	2	14	9	5	7	8	19	6	12	13	9	7	8	7	6	13		12	157

FIP Overlap 2021

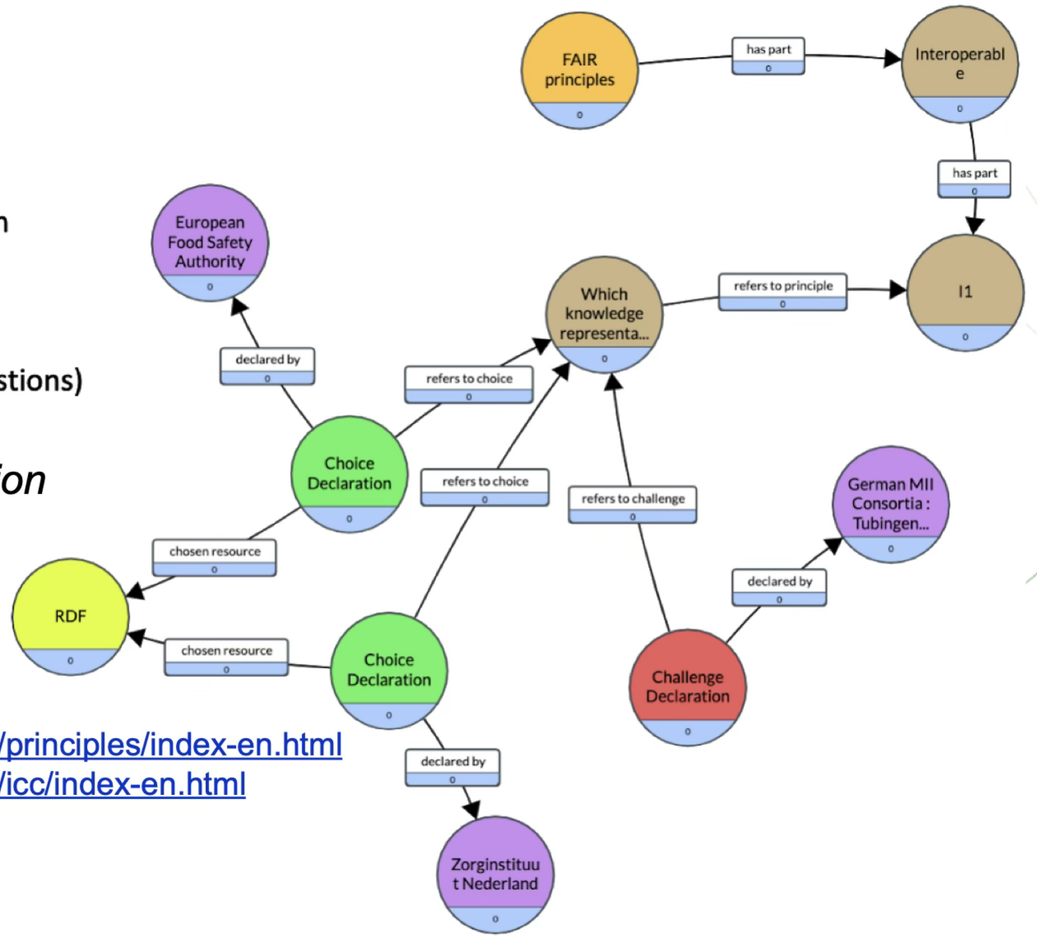
	ACTRIS_DVAS	ACTRIS_GRES	ACTRIS_InSitu	ACTRIS_CLU	ACTRIS-ARES	ACTRIS_ASC	IAGOS	EISCAT	ArgoGdac2	EMSO	lw-marine	SeaDataNet-CDI	SeaDataNet-Sextant	EPOS-ERIC	Anaee	AnaEE_CREA	DANUBIUS	DiSSCo	eLTER-RI	LWERIC_Ecosystem	ICOS	SIOS	
ACTRIS_DVAS		11	11	8	8	11	9	2	6	4	4	2	5	3	7	4	4	3	4	5	7	7	125
ACTRIS_GRES	11		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



Example visualisation: FAIR Data Trains

- FAIR Data Trains
- Resource chosen by more than one Train
- Choice declaration
- Challenge declaration
- FAIR principles concept
- I principles and related choices (i.e. questions)

"Which knowledge representation languages (allowing machine-interoperation) do you use?"

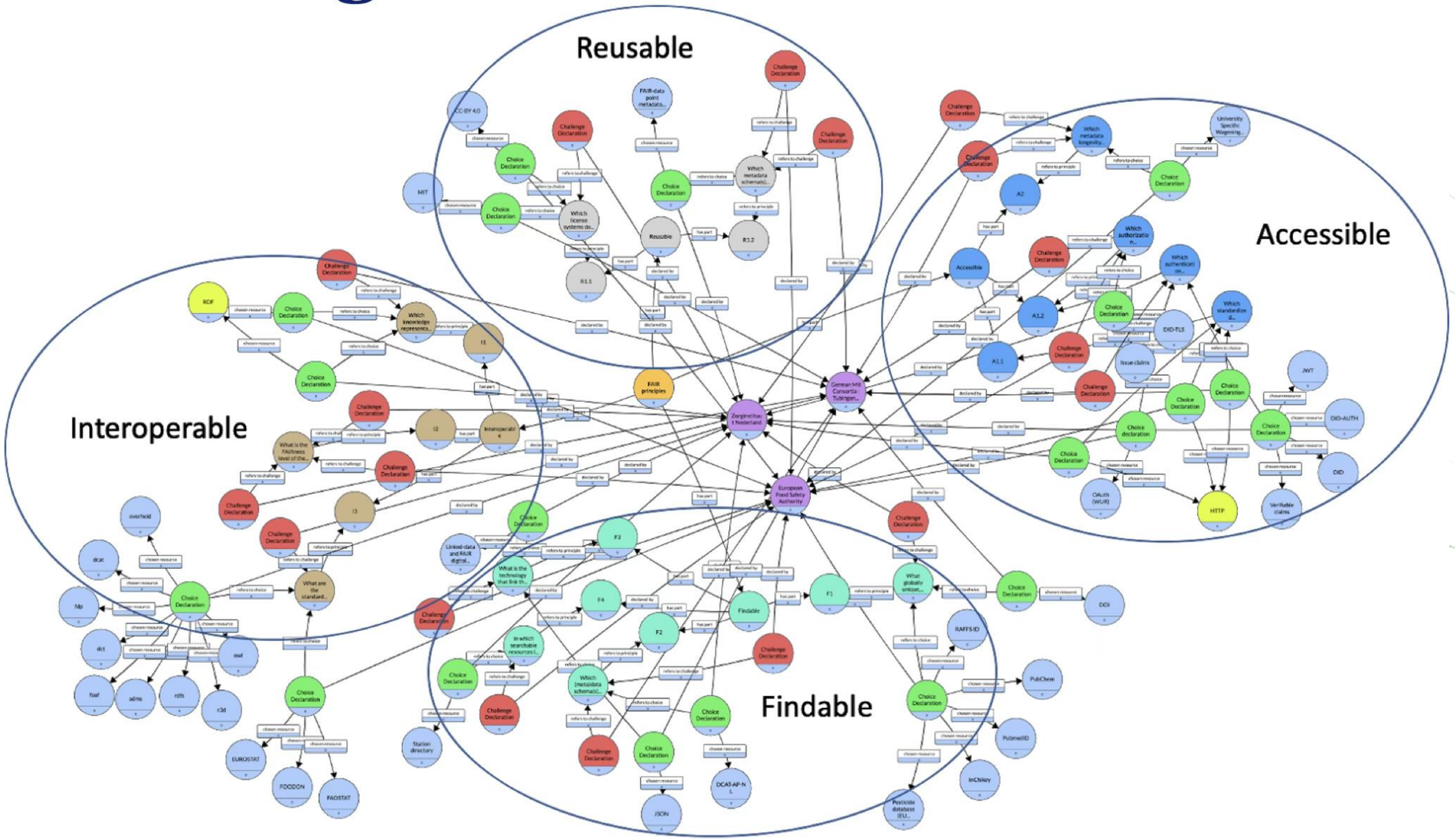


Ontologies used:
https://peta-pico.github.io/FAIR-nanopubs/principles/index-en.html
https://peta-pico.github.io/FAIR-nanopubs/icc/index-en.html

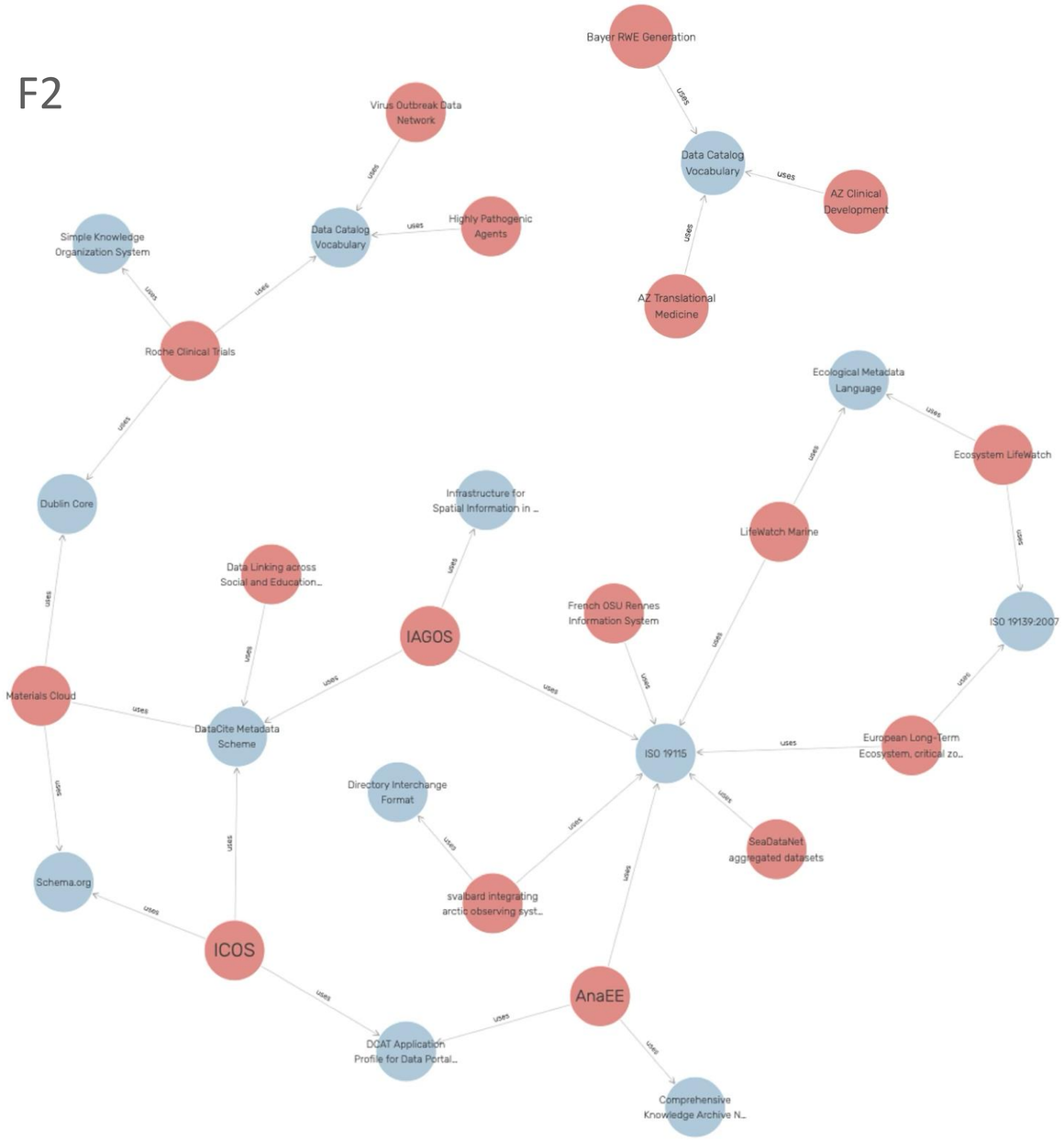
Made with:

GRAFO https://gra.fo

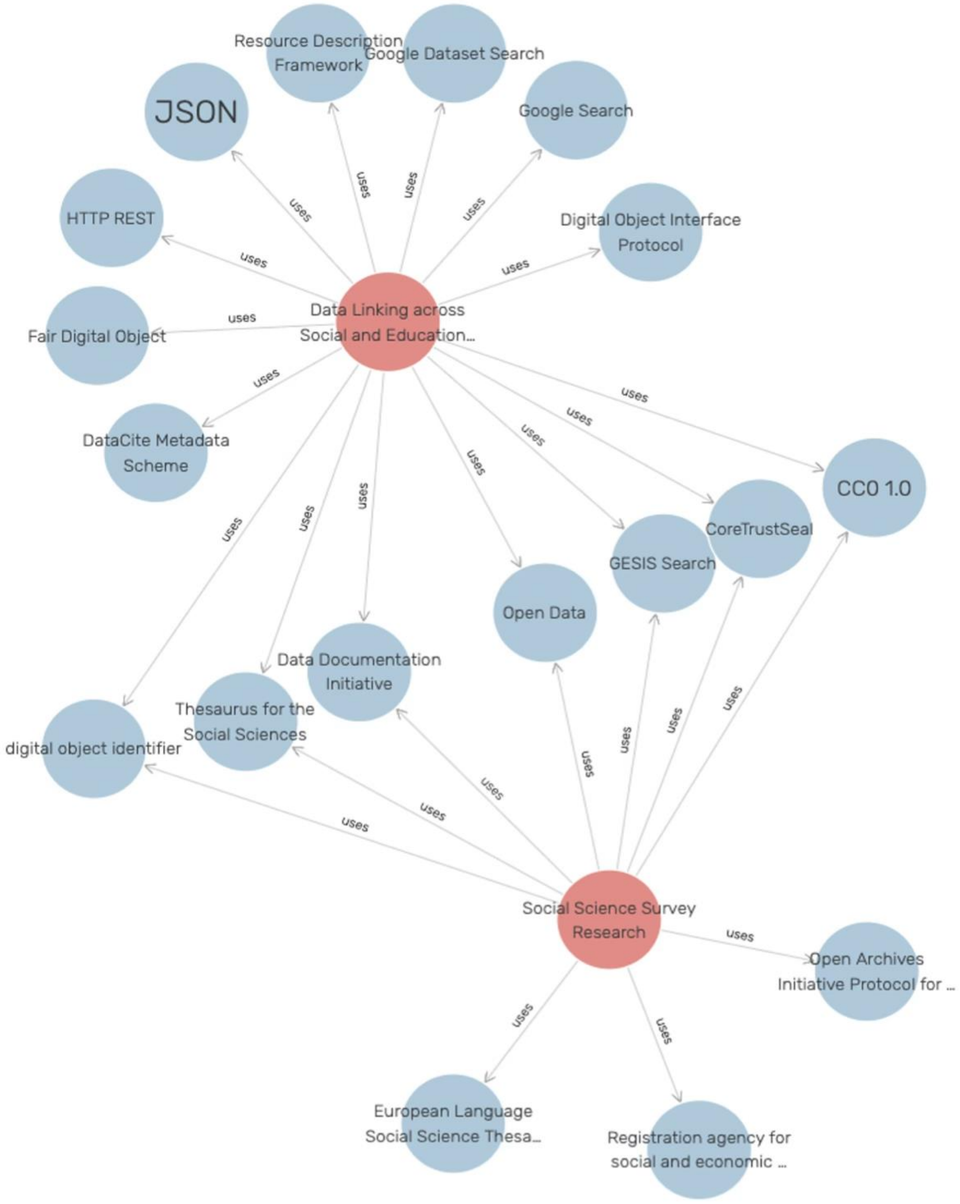
Convergence on FAIR Data Trains

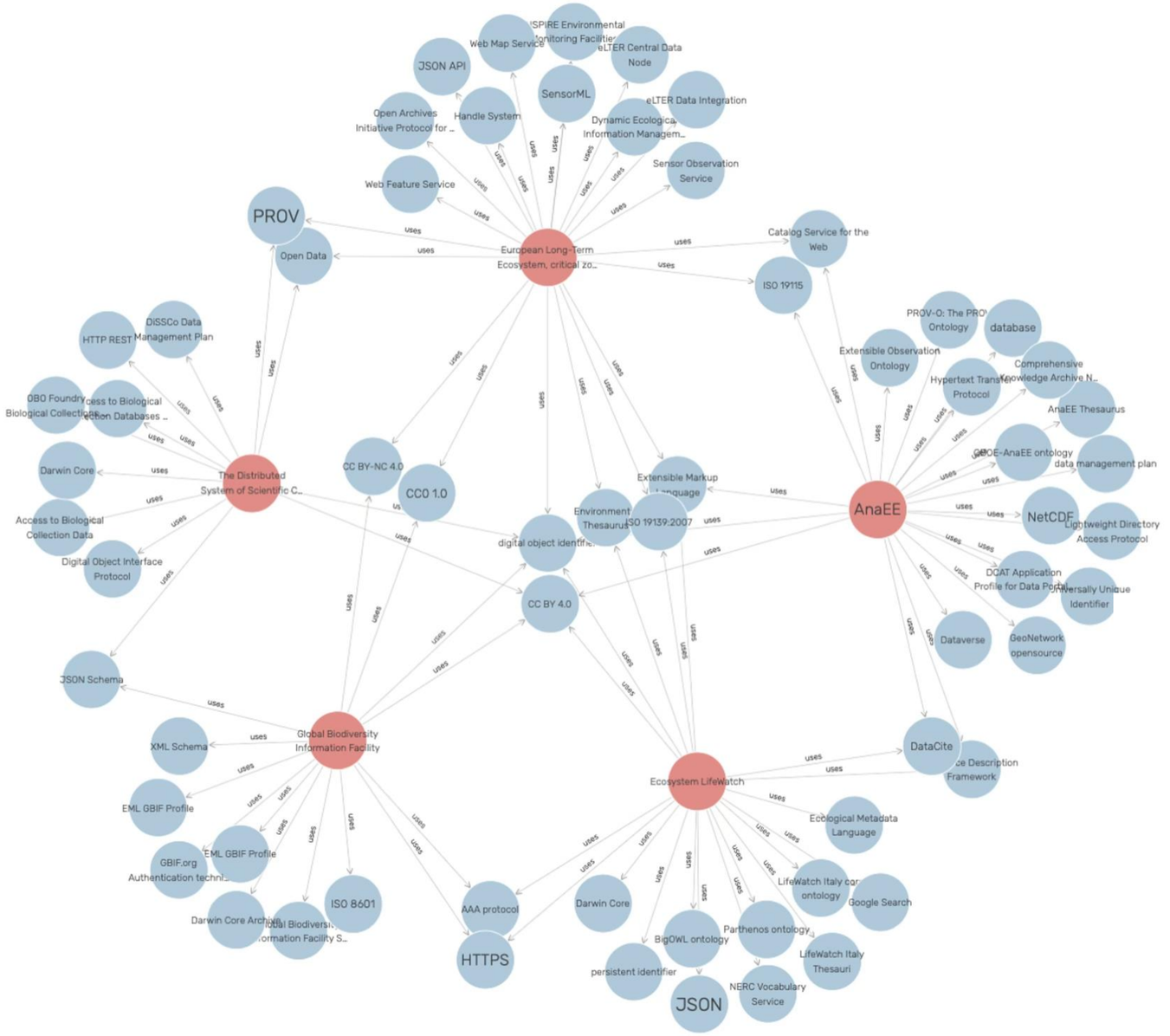


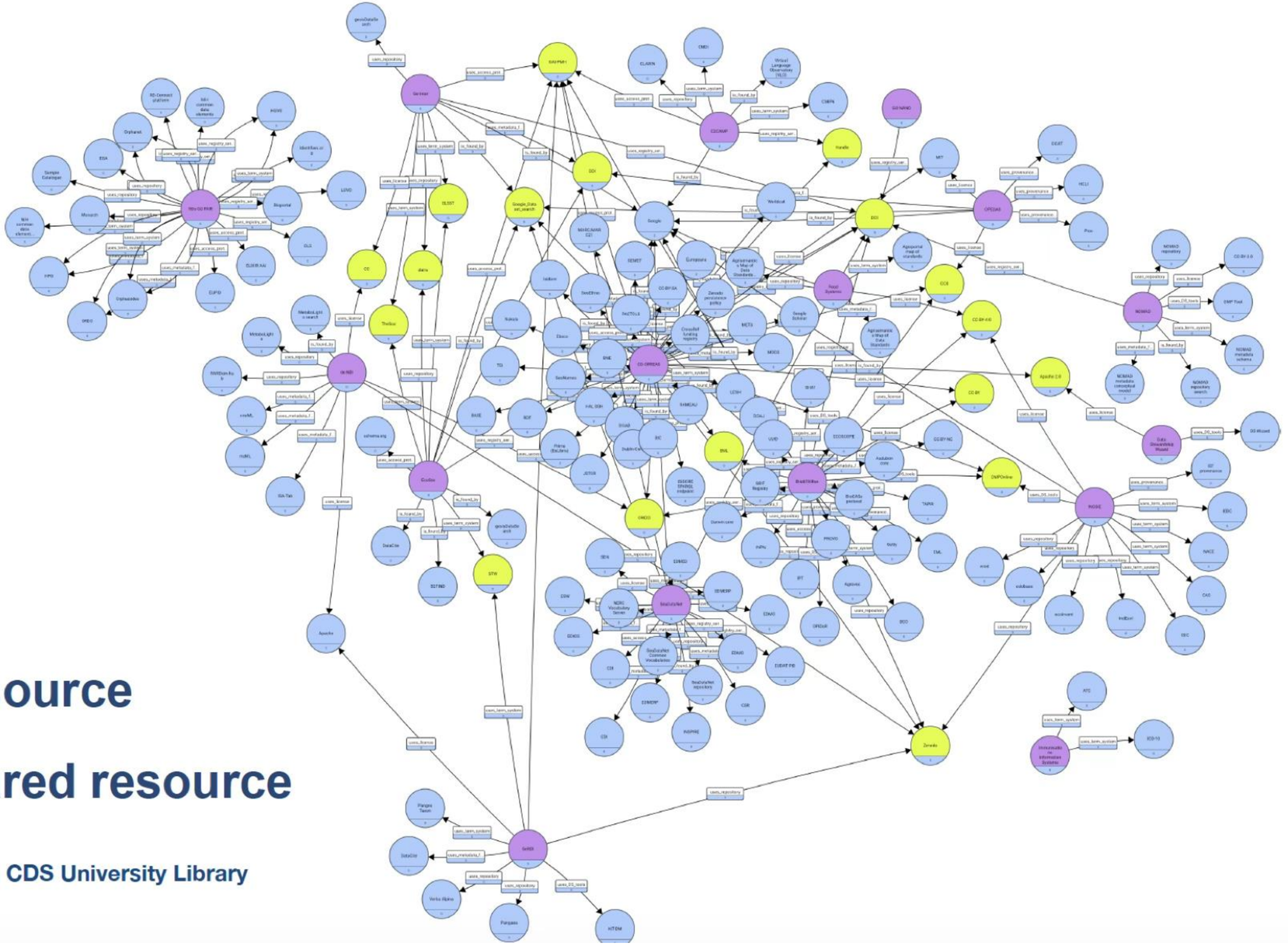
F2


Bayer RWE Generation
Virus Outbreak Data Network
Data Catalog Vocabulary
AZ Clinical Development
Simple Knowledge Organization System
Highly Pathogenic Agents
AZ Translational Medicine
Roche Clinical Trials
Ecological Metadata Language
Ecosystem LifeWatch
Dublin Core
Infrastructure for Spatial Information in ...
LifeWatch Marine
Data Linking across Social and Education...
French OSU Rennes Information System
ISO 19139:2007
IAGOS
Materials Cloud
DataCite Metadata Scheme
ISO 19115
European Long-Term Ecosystem, critical zo...
Directory Interchange Format
SeaDataNet aggregated datasets
Schema.org
svalbard integrating arctic observing syst...
ICOS
AnaEE
DCAT Application Profile for Data Portal...
Comprehensive Knowledge Archive N...
uses


ISO 19115
Hypertext Transfer Protocol
CC BY 4.0
GeoNetwork opensource
French OSU Rennes Information System
OSURIS longevity policy
TAXREF
General Multilingual Environmental Thesa...
Extensible Markup Language
Universally Unique Identifier
uses


Resource Description Framework
Google Dataset Search
JSON
Google Search
HTTP REST
Digital Object Interface Protocol
Data Linking across Social and Education...
Fair Digital Object
DataCite Metadata Scheme
CC0 1.0
CoreTrustSeal
GESIS Search
Open Data
Data Documentation Initiative
Thesaurus for the Social Sciences
digital object identifier
Social Science Survey Research
Open Archives Initiative Protocol for ...
European Language Social Science Thesa...
Registration agency for social and economic ...
uses


JSON API
Web Map Service
SensorML
Open Archives Initiative Protocol for ...
Handle System
Dynamic Ecological Information Managem...
eLTER Data Integration
Web Feature Service
Sensor Observation Service
PROV
Open Data
European Long-Term Ecosystem, critical zo...
Catalog Service for the Web
ISO 19115
HTTP REST
DiSSCo Data Management Plan
OBO Foundry
Darwin Core
Access to Biological Collection Data
Digital Object Interface Protocol
The Distributed System of Scientific C...
CC BY-NC 4.0
CC0 1.0
Extensible Markup Language
Environment Thesaurus
ISO 19139:2007
digital object identifi...
CC BY 4.0
AnaEE
PROV-O: The PROV Ontology
database
Extensible Observation Ontology
Hypertext Transfer Protocol
Comprehensive Knowledge Archive N...
AnaEE Thesaurus
GBOE-AnaEE ontology
data management plan
NetCDF
Lightweight Directory Access Protocol
DCAT Application Profile for Data Portals
Universally Unique Identifier
Dataverse
GeoNetwork opensource
JSON Schema
Global Biodiversity Information Facility
XML Schema
EML GBIF Profile
GBIF.org Authentication techni...
ISO 8601
AAA protocol
HTTPS
Ecosystem LifeWatch
DataCite
Ecological Metadata Language
Darwin Core
BigOWL ontology
Parthenos ontology
LifeWatch Italy Thesauri
Google Search
persistent identifier
JSON
NERC Vocabulary Service

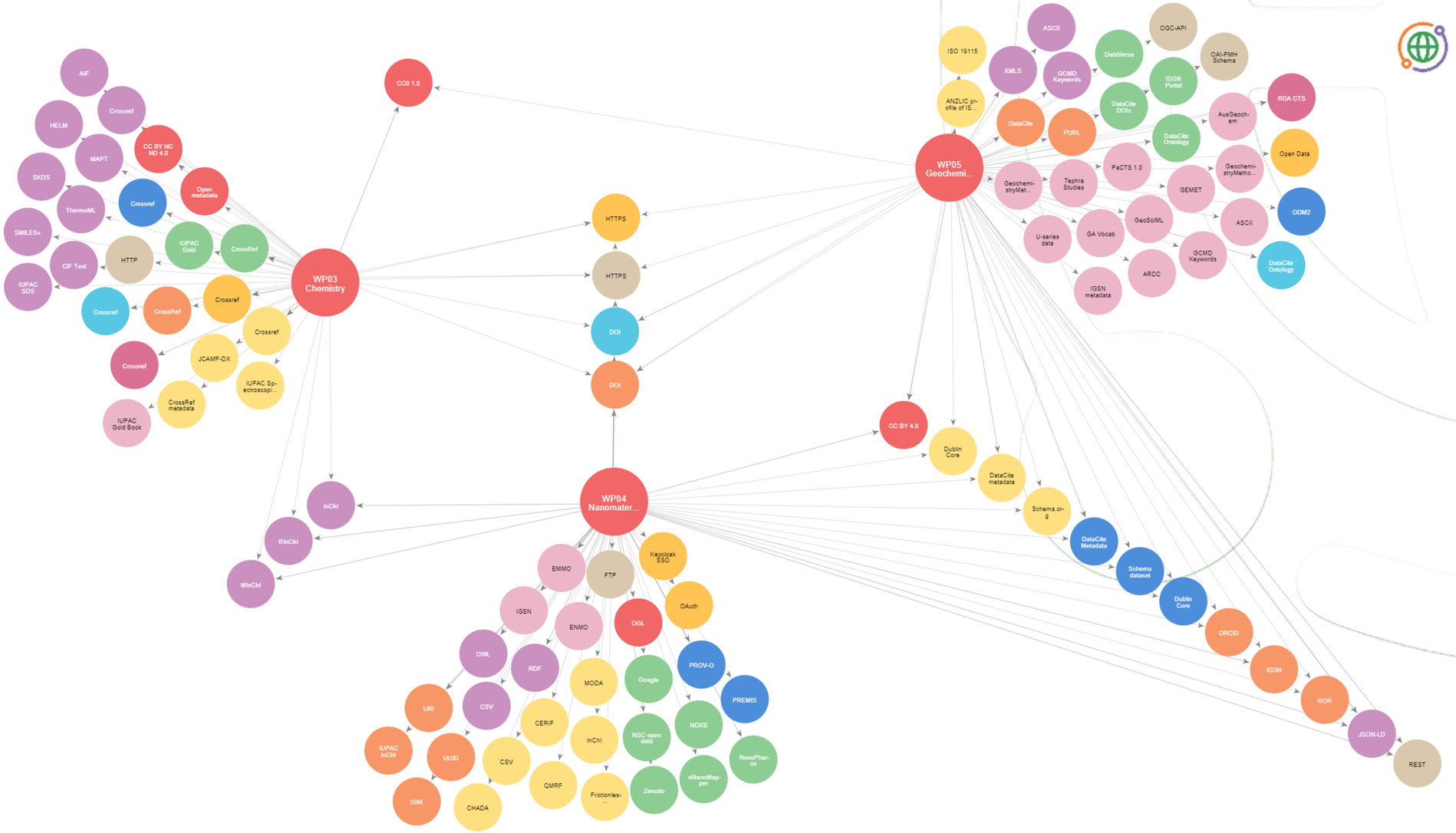


INs

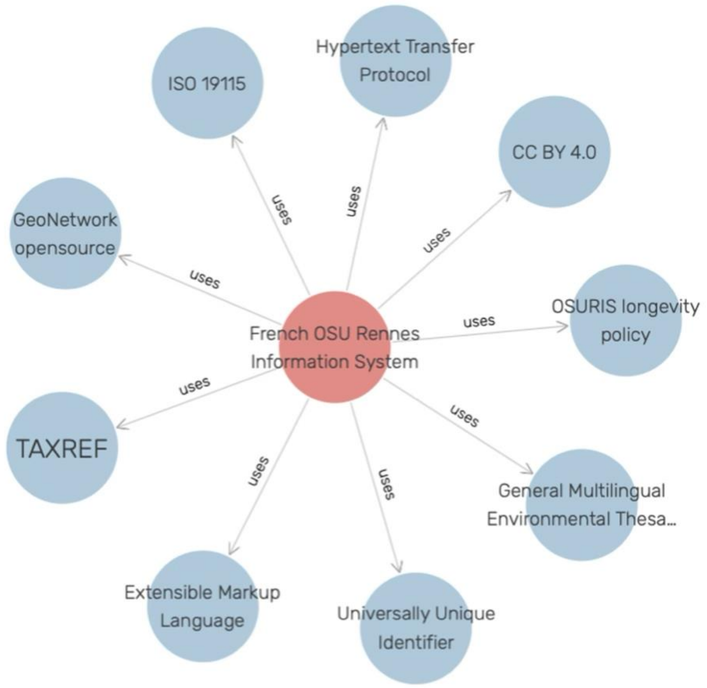
Resource

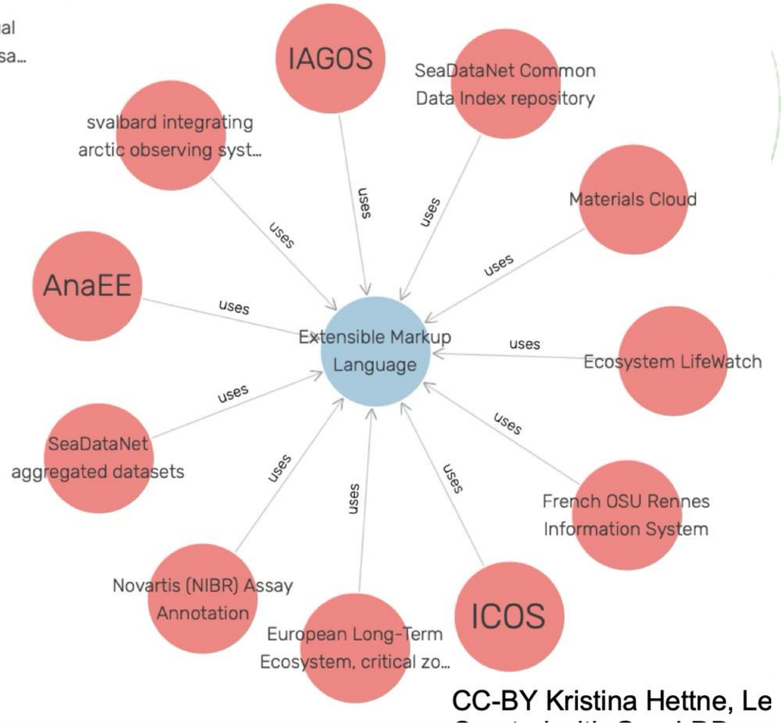
Shared resource

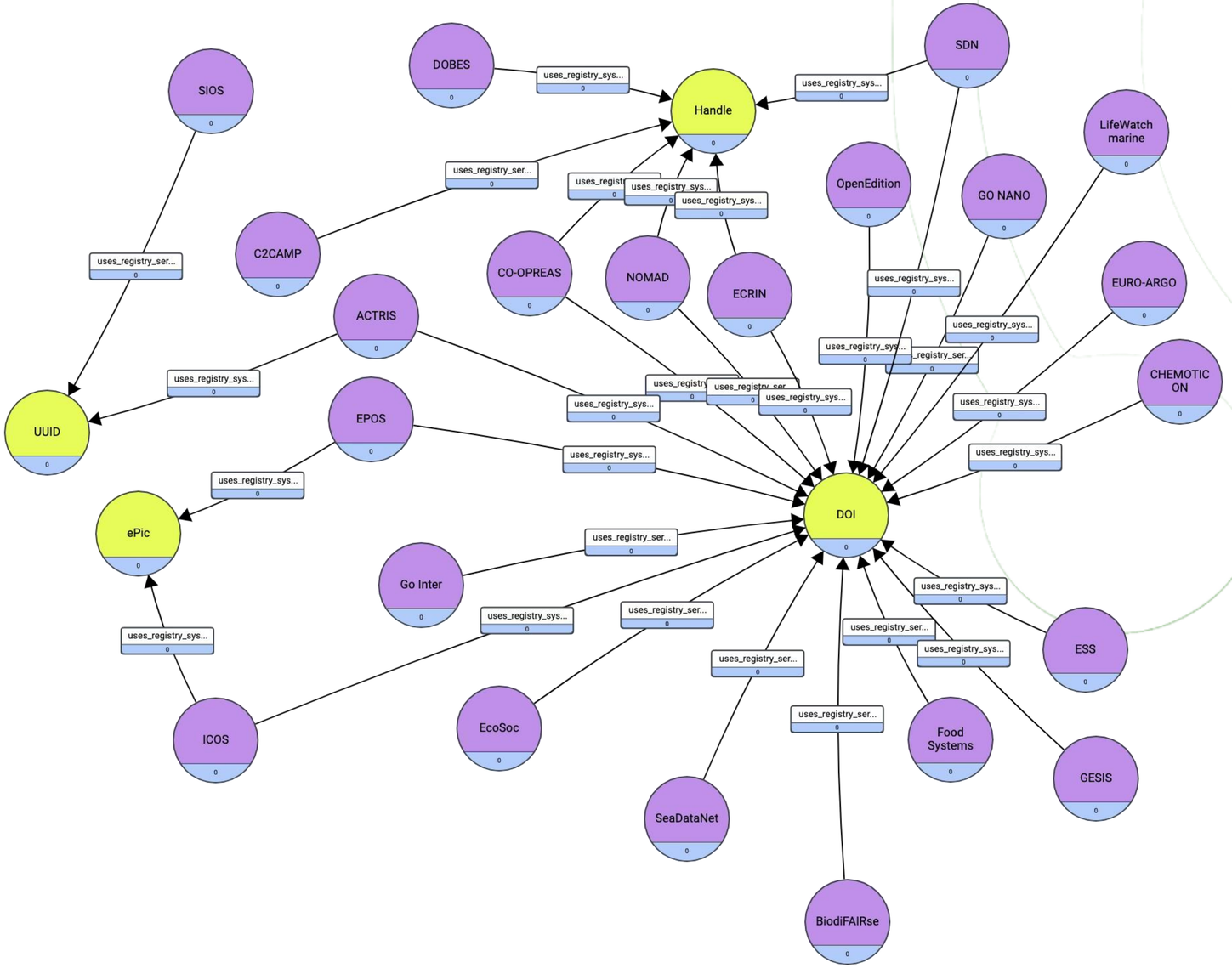
Kristina Hettne, CDS University Library



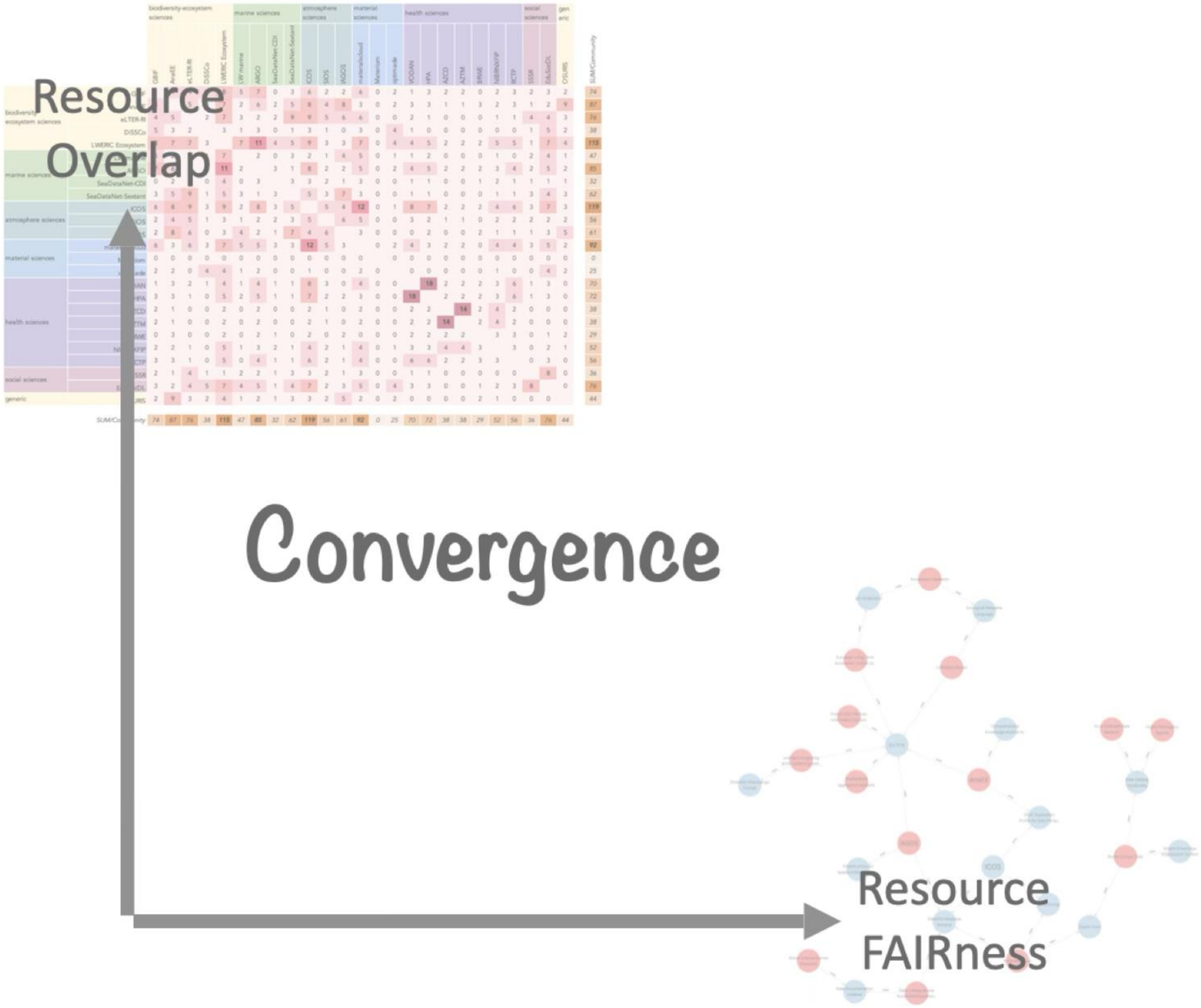
WorldFAIR Project, & Bolland, J. (2024). WorldFAIR FAIR Convergence Graph. Zenodo. https://doi.org/10.5281/zenodo.13608541







Based on FIPs for 32 European Research Infrastructures



FIP analysis: graph visualization

14:55-15:55
(Kristina)