Test des critères FAIR

François Bonnarel, Françoise Genova, Marco Molinaro, Mark Allen

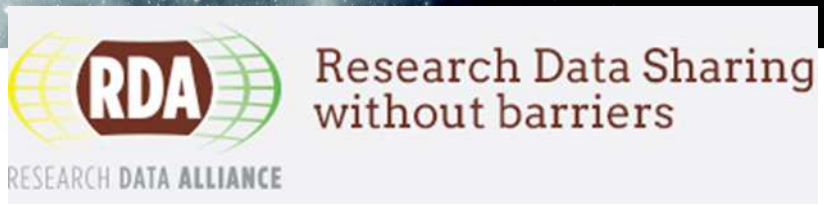
CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

Research Data Alliance activities of interest for ESCAPE – Test of the FAIR criteria

Un travail réalisé dans le cadre du WP4 d'ESCAPE

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064

La Research Data Alliance



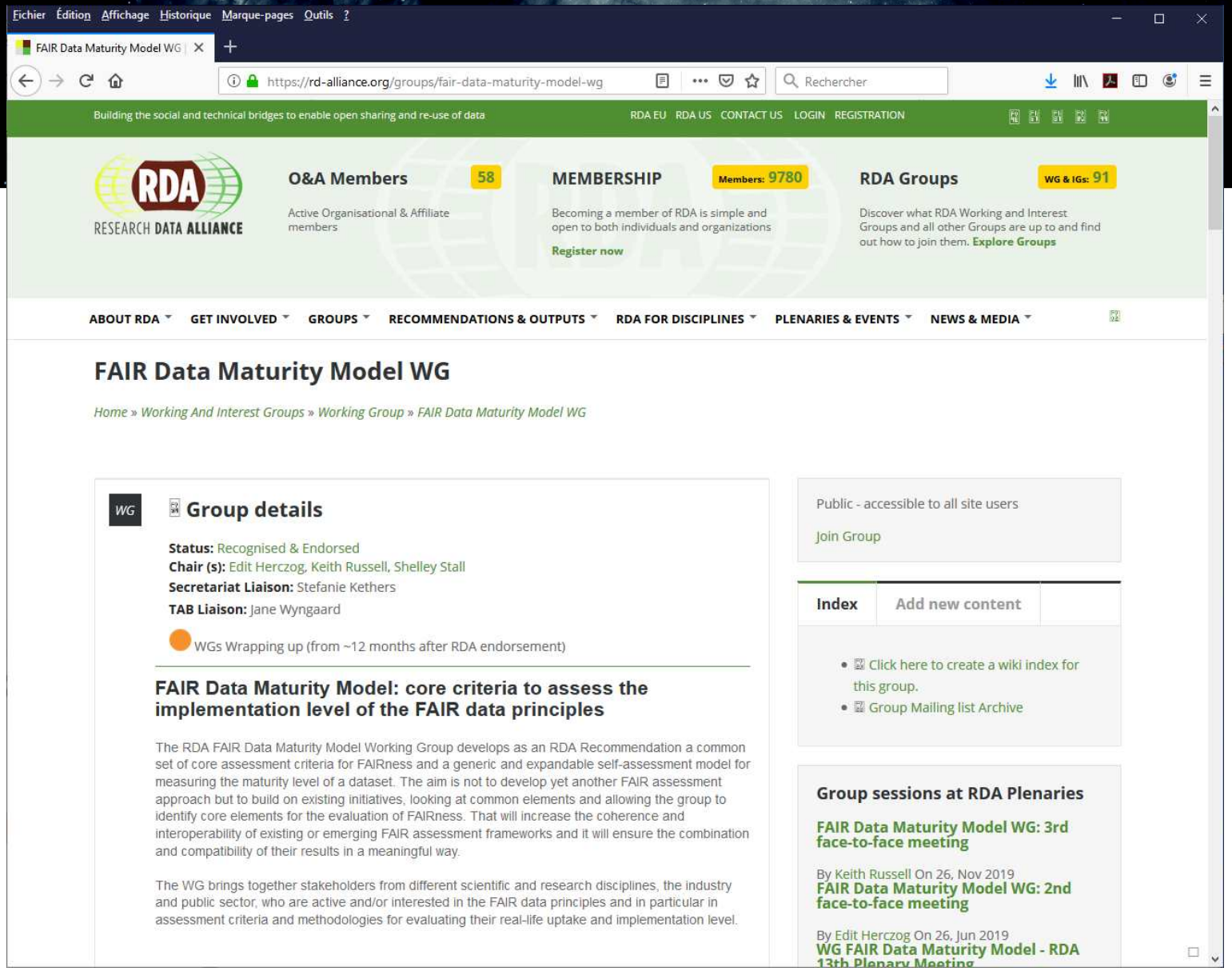
- Créée en 2013 par la Commission Européenne, la NSF (USA) et le Gouvernement Australien
- Un forum neutre, ouvert, international pour discuter tous les aspects du partage des données scientifiques
- La RDA produit des recommandations et des ‘produits de référence
- ~10 000 membres de 144 pays
- 90 Groupes de Travail et Groupes d’Intérêt qui traitent de nombreux sujets technologiques et sociologiques
- La valeur de la RDA pour l’EOSC

https://www.rd-alliance.org/value-research-data-alliance-european-open-science-cloud-eosc

Parmi les activités de la RDA importantes pour l'ASOV/l'astronomie

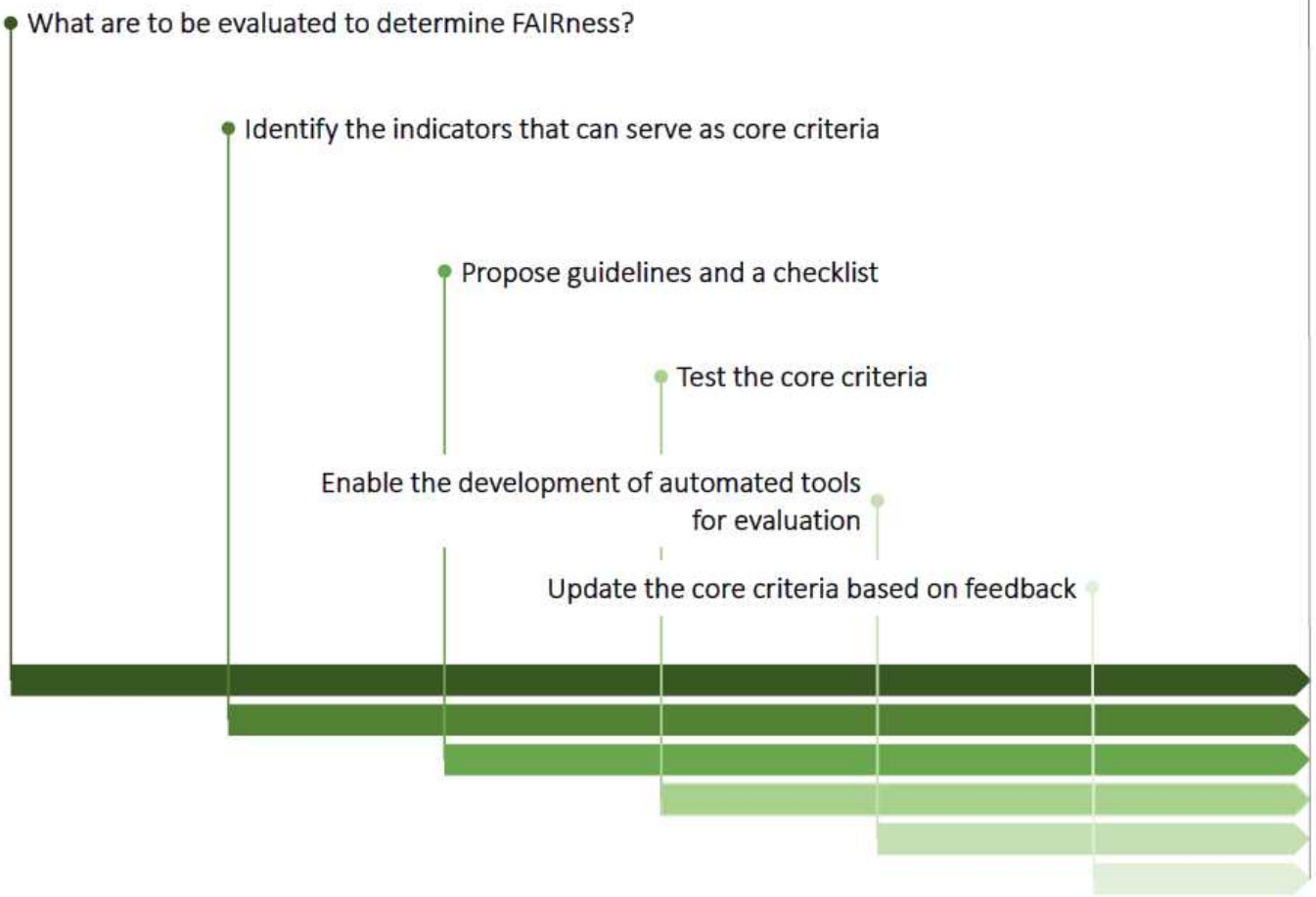
- Au niveau des centres de données: Data Repository Audit and Certification Recommendation >>> Core Trust Seal
  - Atelier OMP en mars 2019
  - « Groupe de Partage » dans le cadre RDA France
  - Ateliers RDA France 12 septembre 2019, 12 mai 2020
  - Atelier « Astronomie » ou « INSU » dans le cadre ASOV?

- Au niveau des données: FAIR Data Maturity Model Working Group
  - Core criteria for FAIR data
  - De très nombreux groupes veulent décider de ce qu'est FAIR. L'objectif du Groupe de Travail est d'avoir une discussion au niveau international sans parti pris


Fichier Édition Affichage Historique Marque-pages Outils ?
FAIR Data Maturity Model WG
https://rd-alliance.org/groups/fair-data-maturity-model-wg
Rechercher
Building the social and technical bridges to enable open sharing and re-use of data
RDA EU RDA US CONTACT US LOGIN REGISTRATION
RESEARCH DATA ALLIANCE
O&A Members 58
Active Organisational & Affiliate members
MEMBERSHIP Members: 9780
Becoming a member of RDA is simple and open to both individuals and organizations
Register now
RDA Groups WG & IGs: 91
Discover what RDA Working and Interest Groups and all other Groups are up to and find out how to join them. Explore Groups
ABOUT RDA GET INVOLVED GROUPS RECOMMENDATIONS & OUTPUTS RDA FOR DISCIPLINES PLENARIES & EVENTS NEWS & MEDIA
FAIR Data Maturity Model WG
Home » Working And Interest Groups » Working Group » FAIR Data Maturity Model WG
WG
Group details
Status: Recognised & Endorsed
Chair (s): Edit Herczog, Keith Russell, Shelley Stall
Secretariat Liaison: Stefanie Kethers
TAB Liaison: Jane Wyngaard
WGs Wrapping up (from ~12 months after RDA endorsement)
FAIR Data Maturity Model: core criteria to assess the implementation level of the FAIR data principles
The RDA FAIR Data Maturity Model Working Group develops as an RDA Recommendation a common set of core assessment criteria for FAIRness and a generic and expandable self-assessment model for measuring the maturity level of a dataset. The aim is not to develop yet another FAIR assessment approach but to build on existing initiatives, looking at common elements and allowing the group to identify core elements for the evaluation of FAIRness. That will increase the coherence and interoperability of existing or emerging FAIR assessment frameworks and it will ensure the combination and compatibility of their results in a meaningful way.
The WG brings together stakeholders from different scientific and research disciplines, the industry and public sector, who are active and/or interested in the FAIR data principles and in particular in assessment criteria and methodologies for evaluating their real-life uptake and implementation level.
Public - accessible to all site users
Join Group
Index
Add new content
Click here to create a wiki index for this group.
Group Mailing list Archive
Group sessions at RDA Plenaries
FAIR Data Maturity Model WG: 3rd face-to-face meeting
By Keith Russell On 26, Nov 2019
FAIR Data Maturity Model WG: 2nd face-to-face meeting
By Edit Herczog On 26, Jun 2019
WG FAIR Data Maturity Model - RDA 13th Plenary Meeting

Objectives

FAIR data maturity model



Les critères en novembre 2019

Versions successives des critères https://docs.google.com/spreadsheets/d/1mkjElFrTBPBH0QViODexNur0xNGhJqau0zkL4w8RRAw/edit#gid=1210743571

	PRINCIPLE	INDICATOR_ID	INDICATORS	PRIORITY
F	F1	F1-01M	Metadata is identified by a persistent identifier	Recommended
	F1	F1-01D	Data is identified by a persistent identifier	Mandatory
	F1	F1-02M	Metadata is identified by a universally unique identifier	Recommended
	F1	F1-02D	Data is identified by a universally unique identifier	Mandatory
	F2	F2-01M	standard	Recommended
	F2	F2-02M	Metadata is provided for the discovery-related elements defined by the RDA Metadata IG, as much as possible and relevant, if no domain/discipline-specific metadata standard is available	Recommended
	F3	F3-01M	Metadata includes the identifier for the data	Mandatory
	F4	F4-01M	Metadata is offered/published/exposed in such a way that it can be harvested and indexed	Recommended
A	A1	A1-01M	Metadata includes information about access conditions	Optional
	A1	A1-01D	Data can be accessed manually (i.e. with human intervention)	Recommended
	A1	A1-02D	Data can be accessed automatically (i.e. by a computer program)	Recommended
	A1	A1-02M	Metadata identifier resolves to a metadata record	Optional
	A1	A1-03D	Data identifier resolves to a digital object	Mandatory
	A1	A1-03M	Metadata is accessed through standardised protocol	Recommended
	A1	A1-04D	Data is accessible through standardised protocol	Recommended
	A1.1	A1.1-01M	Metadata is accessible through a free access protocol	Mandatory
	A1.1	A1.1-01D	Data is accessible through a free access protocol	Mandatory
	A1.1	A1.1-02M	Metadata is accessible through an open-source access protocol	Recommended
	A1.1	A1.1-02D	Data is accessible through an open-source access protocol	Recommended
	A1.1	A1.1-03D	Actions to be taken by a reuser to get access to the data are well documented	Recommended
	A1.2	A1.2-01M	Metadata includes information relevant for access control	Mandatory
	A1.2	A1.2-01D	Data is accessible through an access protocol that supports authentication	Recommended
	A1.2	A1.2-02D	Data is accessible through an access protocol that supports authorisation	Recommended
	A2	A2-01M	Metadata is guaranteed to remain available after data is no longer available	Mandatory
I	I1	I1-01M	Metadata uses knowledge representation expressed in standardised format	Recommended
	I1	I1-01D	Data uses knowledge representation expressed in standardised format	Recommended
	I1	I1-02M	Metadata uses machine-understandable knowledge representation	Optional
	I1	I1-02D	Data uses machine-understandable knowledge representation	Optional
	I1	I1-03M	Metadata uses self-describing knowledge representation	Optional
	I1	I1-03D	Data uses self-describing knowledge representation	Optional
	I2	I2-01M	Metadata uses standard vocabularies	Recommended
	I2	I2-01D	Data uses standard vocabularies	Recommended
	I2	I2-02M	Metadata uses FAIR-compliant vocabularies	Optional
	I2	I2-02D	Data uses FAIR-compliant vocabularies	Optional
	I3	I3-01M	Metadata includes references to other metadata	Recommended
	I3	I3-01D	Data includes references to other data	Recommended
	I3	I3-02M	Metadata includes references to other data	Recommended
	I3	I3-02D	Data includes sufficiently qualified references to other data	Optional
	I3	I3-03M	Metadata includes sufficiently qualified references to other metadata	Recommended
	I3	I3-04M	Metadata include sufficiently qualified references to other data	Optional
R	R1	R1-01M	standard	Recommended
	R1	R1-02M	Metadata is provided for the reuse-related elements defined by the RDA Metadata IG, as much as possible and relevant, if no domain/discipline-specific metadata standard is available	Recommended
	R1.1	R1.1-01M	Metadata includes information about the licence under which the data can be reused	Mandatory
	R1.1	R1.1-02M	Metadata refers to a standard reuse licence	Recommended
	R1.1	R1.1-03M	Metadata includes licence information in the appropriate element of the metadata standard used	Mandatory
	R1.1	R1.1-04M	Metadata refers to a machine-understandable reuse licence	Optional
	R1.1	R1.1-05M	Metadata includes information about consent for reuse (e.g. for personal data)	Recommended
	R1.2	R1.2-01M	Metadata includes provenance information according to community-specific guidelines	Recommended
	R1.2	R1.2-02M	Metadata includes provenance information according to a cross-domain language	Optional
	R1.3	R1.3-01M	Metadata complies with a community standard	Mandatory
	R1.3	R1.3-01D	Data complies with a community standard	Mandatory
	R1.3	R1.3-02M	Metadata is expressed in compliance with a machine-understandable community standard	Optional
	R1.3	R1.3-02D	Data is expressed in compliance with a machine-understandable community standard	Optional

Test par rapport aux pratiques FAIR de l'astronomie

- Les pratiques des différentes disciplines pour le partage des données sont très différentes
- Il est essentiel de tester les critères par rapport aux pratiques disciplinaires
- … en particulier les pratiques de l'astronomie: réutiliser et faire interopérer les données est au coeur de notre démarche scientifique
- Nous avons, avec l'OV, un cadre international de partage des données scientifiques qui permet de trouver les données, d'y accéder, de les faire intreropérer (ainsi que les applications) et de les réutiliser
- Test global en novembre 2019: les critères vs. nos besoins
- Test détaillé, critère par critère, en janvier/février 2020: CDS/VizieR & Trieste VIALACTEA/IA2 → Merci Gilles Landais

Résultats principaux du test global

- Nos objectifs principaux sont la réutilisation et l'interopérabilité, pas la reproductibilité, contrairement à d'autres disciplines
- Nous pouvons en gros vivre avec les critères
- L'OV permet de remplir certains d'entre eux une fois qu'il est implémenté pour donner accès aux données
- Mais nous avons des problèmes avec les priorités données aux critères
  - Find est un process dynamique dans l'OV – recherche possible avec de nombreux paramètres, le PID est seulement l'un d'entre eux
  - Access/réutilisation: au cœur de l'OV mais souvent non conforme. Les données astronomiques sont en général ouvertes et n'ont pas toujours de licence d'utilisation – réutilisation en suivant l'éthique disciplinaire (citer ce qu'on utilise). Les données astronomiques sont très largement réutilisées, donc ça n'est pas un obstacle
  - En novembre, pas de critère essentiel pour l'intéropérabilité, depuis plusieurs ont été ajoutés. Cela montre le caractère un peu ad hoc des priorités sur les critères
- Certaines métadonnées sont au niveau des données, d'autres au niveau de la collection

Conclusions of the overall test

- Real critical problem with mandatory (or essential) criteria – we would chose different ones!
- FAIRness is a process: the evaluation method should be inclusive and allow and measure progress (compliance scales)
- Open by default should be considered as acceptable
- There is a cost to adapt a repository to fulfill criteria different from the disciplinary ones, and archives are not supported for that
- Detailed test ongoing

Comparison of FAIR priority criteria

	PRINCIPLE	INDICATOR_ID	INDICATORS	PRIORITY
F	F1	F1-01M	Metadata is identified by a persistent identifier	Essential
	F1	F1-01D	Data is identified by a persistent identifier	Essential
	F1	F1-02M	Metadata is identified by a universally unique identifier	Essential
	F1	F1-02D	Data is identified by a universally unique identifier	Essential
	F2	F2-01M	Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard	Important
	F3	F3-01M	Metadata includes the identifier for the data	Important
	F4	F4-01M	Metadata is offered/published/exposed in such a way that it can be harvested and indexed	Essential
A	A1	A1-01M	Metadata includes information about access conditions	Essential
	A1	A1-02D	Data can be accessed manually (i.e. with human intervention)	Essential
	A1	A1-02D	Data can be accessed automatically (i.e. by a computer program)	Important
	A1	A1-02M	Metadata identifier resolves to a metadata record	Essential
	A1	A1-03D	Data identifier resolves to a digital object	Essential
	A1	A1-03M	Metadata is accessed through standardised protocol	Important
	A1	A1-04D	Data is accessible through standardised protocol	Important
	A1.1	A1.1-01M	Metadata is accessible through a free access protocol	Essential
	A1.1	A1.1-02D	Data is accessible through a free access protocol	Important
	A1.1	A1.1-02M	Metadata is accessible through an open-source access protocol	Important
	A1.1	A1.1-03D	Data is accessible through an open-source access protocol	Important
	A1.1	A1.1-03D	Actions to be taken by a reuser to get access to the data are well documented	Important
	A1.2	A1.2-01M	Metadata includes information relevant for access control	Essential
	A1.2	A1.2-01D	Data is accessible through an access protocol that supports authentication	Useful
	A1.2	A1.2-02D	Data is accessible through an access protocol that supports authorisation	Useful
	A2	A2-01M	Metadata is guaranteed to remain available after data is no longer available	Essential
I	I1	I1-01M	Metadata uses knowledge representation expressed in standardised format	Essential
	I1	I1-01D	Data uses knowledge representation expressed in standardised format	Important
	I1	I1-02M	Metadata uses machine-understandable knowledge representation	Essential
	I1	I1-02D	Data uses machine-understandable knowledge representation	Important
	I1	I1-02M	Metadata uses self-describing knowledge representation	Useful
	I1	I1-03D	Data uses self-describing knowledge representation	Useful
	I2	I2-01M	Metadata uses standard vocabularies	Important
	I2	I2-01D	Data uses standard vocabularies	Important
	I2	I2-02M	Metadata uses FAIR-compliant vocabularies	Important
	I2	I2-02D	Data uses FAIR-compliant vocabularies	Useful
	I3	I3-01M	Metadata includes references to other metadata	Important
	I3	I3-01D	Data includes references to other data	Important
	I3	I3-02M	Metadata includes references to other data	Useful
	I3	I3-02D	Data includes sufficiently qualified references to other data	Useful
	I3	I3-03M	Metadata includes sufficiently qualified references to other metadata	Important
	I3	I3-04M	Metadata include sufficiently qualified references to other data	Useful
R	R1	R1-01M	Sufficient metadata is provided to allow reuse, following domain/discipline-specific metadata standard	Essential
	R1.1	R1.1-01M	Metadata includes information about the licence under which the data can be reused	Essential
	R1.1	R1.1-02M	Metadata refers to a standard reuse licence	Important
	R1.1	R1.1-03M	Metadata includes licence information in the appropriate element of the metadata standard used	Essential
	R1.1	R1.1-04M	Metadata refers to a machine-understandable reuse licence	Important
	R1.1	R1.1-05M	Metadata includes information about consent for reuse (e.g. for personal data)	Important
	R1.2	R1.2-01M	Metadata includes provenance information according to community-specific standards	Important
	R1.2	R1.2-02M	Metadata includes provenance information according to a cross-domain language	Useful
	R1.3	R1.3-01M	Metadata complies with a community standard	Important
	R1.3	R1.3-01D	Data complies with a community standard	Important
	R1.3	R1.3-02M	Metadata is expressed in compliance with a machine-understandable community standard	Important
	R1.3	R1.3-02D	Data is expressed in compliance with a machine-understandable community standard	Important

d column colour code:
een: agree that mandatory
ue: mandatory for astronomy,
ot for the RDA WG
d: mandatory for RDA WG,
ot for astronomy

Funded by the European Union's Horizon 2020 - Grant N° 824064

Conclusions

- Un problème réel avec les critères ‘mandatory’/’essential’
- FAIR est un process: la méthode d’évaluation devrait être inclusive, identifier les domaines où il faut progresser et mesurer les progrès (échelle de conformité plutôt que oui/non) – semble être accepté
- Il faudrait qu’ouvert par défaut soit possible – ne sera probablement pas accepté, les juristes insistent sur la nécessité d’une licence pour que les données puissent être réutilisées
- Discussion sur les licences dans l’IVOA (IG Data Curation & Préservation)? Permettrait d’avoir la participation des grands producteurs de données

IVOA Note being written

FAIR practices in Astronomy, and how they fit with the FAIR criteria

Status report

Françoise Genova, François Bonnarel (CDS/Observatoire Astronomique de Strasbourg), Marco Molinaro (INAF), Mark Allen (CDS/Observatoire Astronomique de Strasbourg)

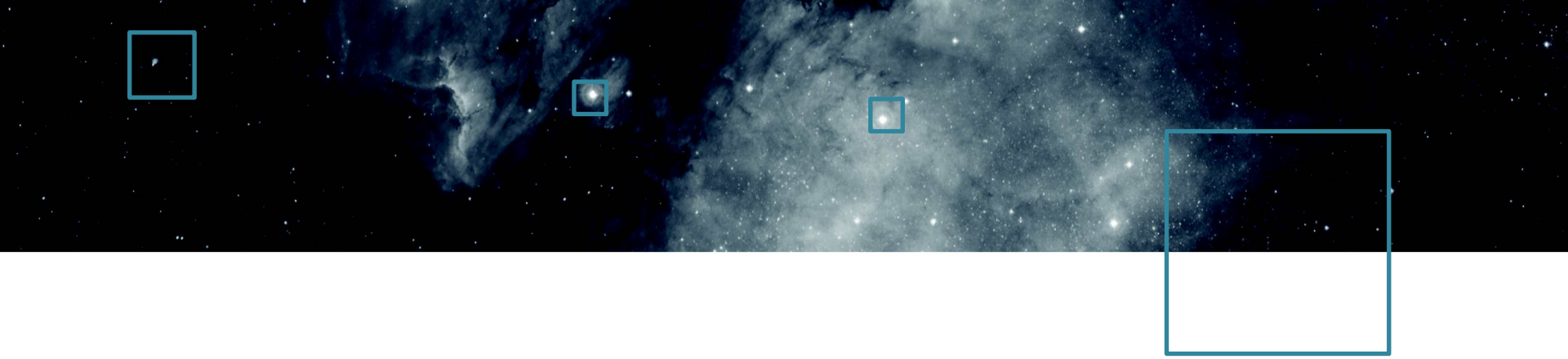
ESCAPE WP4 - Connecting ESFRI projects to EOSC through VO framework (CEVO)

V0.4, 30 January 2020

1. Introduction

Disciplinary data sharing frameworks are the pillars of open science. A key element for them to succeed is to be driven by the community science needs, and to take into account community practices (e.g. Genova et al., 2017[1]). Astronomy has been a pioneer of Open Data Sharing, and remains at the forefront. International agreement on a data format, FITS, and on standards allowing users to find, access and interoperate data, the so-called Virtual Observatory (VO) standards, provides a shared data sharing framework open to all and enables the development of interoperable tools to access and use data.

FITS was first published in 1981 (Wells, Greisen & Harten[2]). The International Virtual Observatory Alliance[3] (IVOA) was created in 2002 to define and maintain the astronomical interoperability standards. Its processes[4] are adapted from the W3C ones. Data producers have been providing their data in the Virtual Observatory, and the community has been able to find, access, interoperate and reuse astronomical data years before FAIR was defined in the foundational Nature paper in 2016 (Wilkinson et al[5]). More than 100 "authorities" from all around the world, large agencies as well as smaller teams, have declared at least a service in the VO registry of resources. VO-enabled data services provided by the ground and space-based observatory archives and value-added data repositories, as well as the VO-enabled interoperable tools providing access to data, are used by the community in their daily research work.

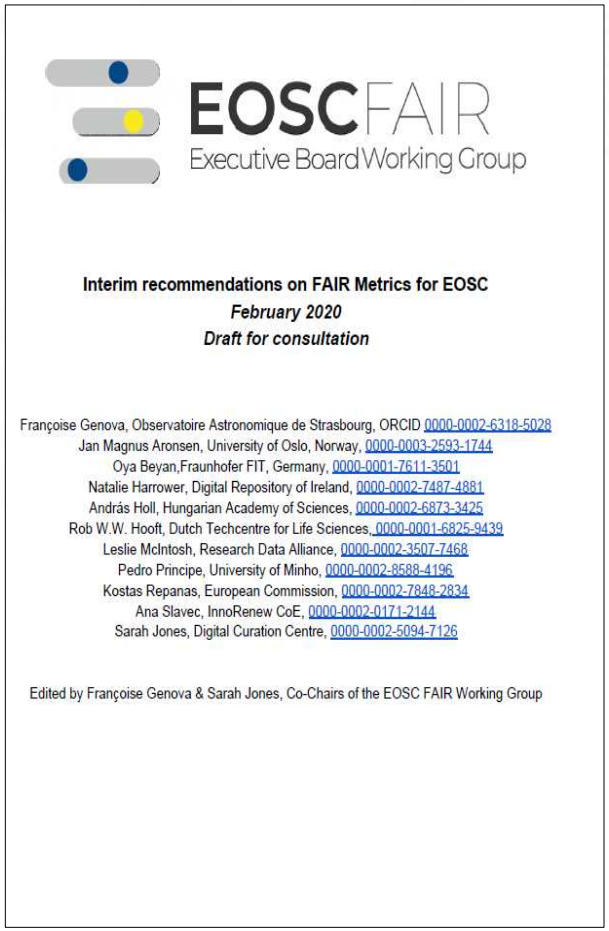


AU NIVEAU EUROPÉEN

EOSC FAIR Working Group

- Recommendations on the implementation of Open and FAIR practices within the EOSC.
- FAIR work plan based on “Turning FAIR into Reality” and work of key projects like FAIRsFAIR, RDA and FREYA
- Chair: Sarah Jones (DCC), Vice-chair: Françoise Genova
- 4 Task Forces, 5 outputs
  - Investigation and report on FAIR practices
  - A Persistent Identifier (PID) policy for EOSC (with the Architecture WG)
  - FAIR metrics assessing datasets and other digital objects
  - Repository certification guidelines for supporting FAIR research outputs
  - An EOSC interoperability framework

Interim recommendations open for comments

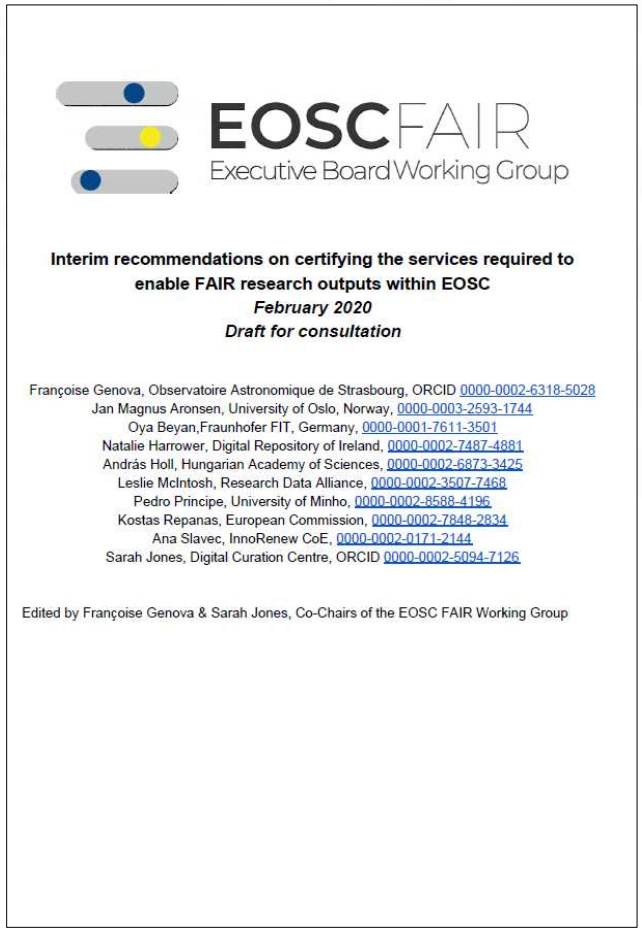


EOSCFAIR
Executive Board Working Group

Interim recommendations on FAIR Metrics for EOSC
February 2020
Draft for consultation

Françoise Genova, Observatoire Astronomique de Strasbourg, ORCID 0000-0002-6318-5028
Jan Magnus Aronsen, University of Oslo, Norway, 0000-0003-2593-1744
Oya Beyan,Fraunhofer FIT, Germany, 0000-0001-7611-3501
Natalie Harrower, Digital Repository of Ireland, 0000-0002-7487-4881
András Holl, Hungarian Academy of Sciences, 0000-0002-6873-3425
Rob W.W. Hooft, Dutch Techcentre for Life Sciences, 0000-0001-6825-9439
Leslie McIntosh, Research Data Alliance, 0000-0002-3507-7468
Pedro Principe, University of Minho, 0000-0002-8588-4196
Kostas Repanas, European Commission, 0000-0002-7848-2834
Ana Slavec, InnoRenew CoE, 0000-0002-0171-2144
Sarah Jones, Digital Curation Centre, 0000-0002-5094-7126

Edited by Françoise Genova & Sarah Jones, Co-Chairs of the EOSC FAIR Working Group



EOSCFAIR
Executive Board Working Group

Interim recommendations on certifying the services required to enable FAIR research outputs within EOSC
February 2020
Draft for consultation

Françoise Genova, Observatoire Astronomique de Strasbourg, ORCID 0000-0002-6318-5028
Jan Magnus Aronsen, University of Oslo, Norway, 0000-0003-2593-1744
Oya Beyan,Fraunhofer FIT, Germany, 0000-0001-7611-3501
Natalie Harrower, Digital Repository of Ireland, 0000-0002-7487-4881
András Holl, Hungarian Academy of Sciences, 0000-0002-6873-3425
Leslie McIntosh, Research Data Alliance, 0000-0002-3507-7468
Pedro Principe, University of Minho, 0000-0002-8588-4196
Kostas Repanas, European Commission, 0000-0002-7848-2834
Ana Slavec, InnoRenew CoE, 0000-0002-0171-2144
Sarah Jones, Digital Curation Centre, ORCID 0000-0002-5094-7126

Edited by Françoise Genova & Sarah Jones, Co-Chairs of the EOSC FAIR Working Group

https://www.eoscsecretariat.eu/eosc-liaison-platform/post/interim-recommendations-fair-metrics-and-service-certification-apply