

FAIR Data Action Plan

Interim recommendations and actions from the European Commission
Expert Group on FAIR data

June 2018

Sandra Collins, National Library of Ireland
Françoise Genova, Observatoire Astronomique de Strasbourg
Natalie Harrower, Digital Repository of Ireland
Simon Hodson, CODATA, Chair of the Group
Sarah Jones, Digital Curation Centre, Rapporteur
Leif Laaksonen, CSC-IT Center for Science
Daniel Mietchen, Data Science Institute, University of Virginia
Rūta Petrauskaitė, Vytautas Magnus University
Peter Wittenburg, Max Planck Computing and Data Facility

Disclaimer

The Expert Group operates in full autonomy and transparency. The views and recommendations in this report are those of the Expert Group members acting in their personal capacities and do not necessarily represent the opinions of the European Commission or any other body; nor do they commit the Commission to implement them.

Please cite as: Hodson, Jones et al. (2018) *FAIR Data Action Plan*. Interim recommendations and actions from the European Commission Expert Group on FAIR data <https://doi.org/10.5281/zenodo.1285290>

Preface from the Expert Group

It is recognised that FAIR data (data that are Findable, Accessible, Interoperable and Reusable) play an essential role in the objectives of Open Science to improve and accelerate scientific research, to increase the engagement of society, and to contribute significantly to economic growth. Accordingly, ‘the Open Science agenda contains the ambition to make FAIR data sharing the default for scientific research by 2020.’ The overall objective of the European Commission Expert Group on Turning FAIR data into reality is to help operationalise and facilitate the achievement of this goal.

To this end, this report that examines the FAIR data principles, considers other supporting concepts and discusses the changes necessary, as well as existing activities and stakeholders to make these interventions. Recommendations and actions are presented as an Action Plan for consideration by the Commission, Member States and leading stakeholders in the research and data communities.

It might have been possible to take a data centric point of view and to work through the FAIR principles slavishly or systematically (depending on your point of view) asking what needs to be done to achieve each one. The Expert Group decided at an early point that this would not be the most effective approach to our task. Rather we felt it was important to take a holistic and systemic approach and to describe the broader range of changes required to achieve FAIR data. We hope that what has emerged will be at one and the same time an Action Plan that will be immediately useful and a longer standing survey and discussion, providing a discursive framework for ongoing considerations of how to make FAIR data a reality.

Just as this is interim report, so this is an interim preface. At this stage we are in particular looking for constructive feedback. Does the Action Plan highlight the correct priorities? Are the recommendations sound and the actions tangible and achievable? Are they presented in a way that will helpfully guide the stakeholders mentioned? Is the Action Plan sufficiently grounded in the discussions and arguments of the broader report? Given the way this particularly piece of marble has already been cut and carved, what still needs to be done to make a polished statue emerge?

Consultation on the interim report will be launched at the EOOSC summit on 11 June 2018 and initiated by means of a workshop at that meeting. It will be pursued by online means and by webinars until 5 August. A final version of the Report and Action Plan will be published at the Austrian Presidency event on 23 November.

The group has conducted its work by means of face-to-face and virtual meetings and a lot of asynchronous, collaborative work with the text. All members of the group have contributed substantively and substantially to the text. We hope that we have harnessed the strength and collective wisdom of the Expert Group, while minimising the flaws of group authorship. Our approach has been discursive and we have endeavoured to explore the arguments relating to FAIR in detail to identify the key steps needed for implementation. This is an iterative process and the final version of the report will present a more condensed argument.

The group has been chaired by Simon Hodson, with Sarah Jones as Rapporteur; but in effect the two have acted as co-chairs.

Stakeholder groups assigned Actions

1. **Research communities:** practitioners from all fields of humanities and science, clustered in groups around disciplinary interests, data types or cross-cutting grand challenges.
2. **Data services:** domain repositories, Research Infrastructures (ESFRIs) and E-Infrastructures, institutional provision, community and commercial tools and services.
3. **Data stewards:** support staff from research communities and research libraries, and those managing data repositories.
4. **Standards bodies:** formal organisations and consortia coordinating data standards and governing procedures relevant to FAIR, e.g. repository certification, curriculum accreditation.
5. **Global coordination fora:** the Research Data Alliance, CODATA, WDS Communities of Excellence, FORCE11, GO FAIR and other similar initiatives.
6. **Policymakers:** governments, international entities like OECD, research funders, institutions, publishers and others defining data policy.
7. **Research funders:** the European Commission, national research funders, charitable organisations and foundations, and other funders of research activity.
8. **Institutions:** universities and research performing organisations
9. **Publishers:** commercial and not-for-profit, paywall and Open Access publishers of research papers and data.

Primary Recommendations and Actions

Step 1: Define and apply FAIR appropriately

Rec. 1: Definitions of FAIR

FAIR is not limited to its four constituent elements: it must also comprise appropriate openness, the assessability of data, long-term stewardship, and other relevant features. To make FAIR data a reality, it is necessary to incorporate these concepts into the definition of FAIR.

- The FAIR principles should be consulted on and clarified to ensure they are understood to include appropriate openness, timeliness of sharing, assessability, data appraisal, long-term stewardship and legal interoperability.
Stakeholders: Global coordination fora; Research communities; Data services.
- The term FAIR data is widely-used and effective so should not be extended with additional letters.
Stakeholders: Research communities; Data services.
- The relationship between FAIR and Open should be clarified and well-articulated. FAIR depends on appropriate Openness which can be expressed as ‘as Open as possible, as closed as necessary’.
Stakeholders: Research communities

Related recommendations: [Rec. 2: Mandates and boundaries of Open](#); [Rec. 7: Disciplinary interoperability frameworks](#).

Rec. 2: Mandates and boundaries for Open

The Open Data mandate for publicly funded research should be made explicit in all policy. It is important that the maxim 'as Open as possible, as closed as necessary' be applied proportionately with genuine best efforts to share.

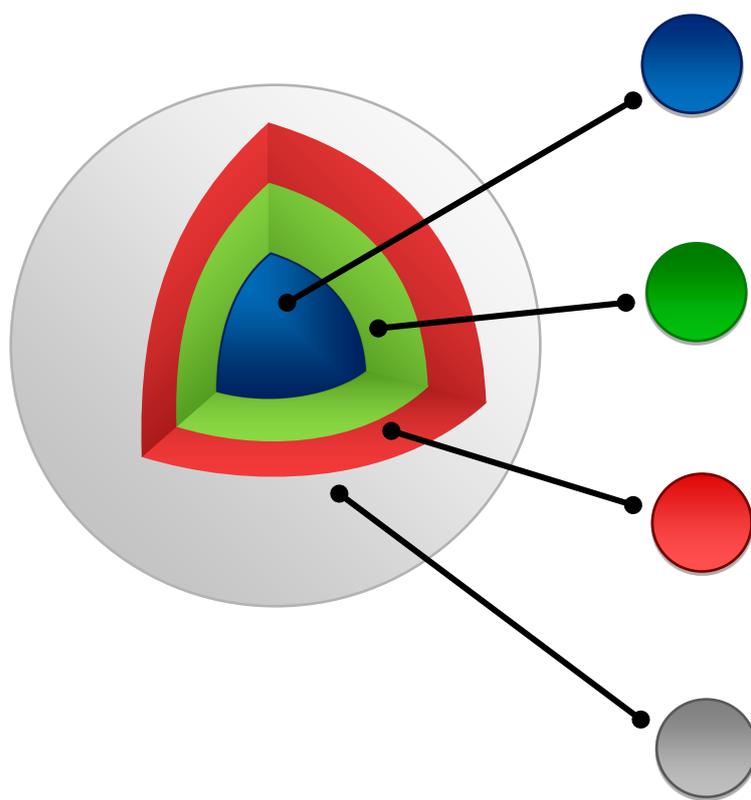
- Steps should be taken to ensure coherence across data policy and issue collective statements of intent wherever possible.
Stakeholders: Research funders; Policymakers.
- Policies should require an explicit and justified statement when data cannot be Open and a proportionate and discriminating course of action to ensure maximum appropriate data accessibility, rather than allowing a wholesale opt out from the mandate for Open Data.
Stakeholders: Funders; Policymakers.
- Sustained work is needed to clarify in more detail the appropriate boundaries of Open, the proportionate exceptions to data sharing and robust processes for data that needs to be protected.
Stakeholders: Research communities; Data services; Global coordination fora.
- Concrete and accessible guidance should be provided to researchers in relation to sharing sensitive and commercial data as openly as possible.
Stakeholders: Data stewards; Data services; Institutions; Publishers.

Related recommendations: [Rec 1: Definitions of FAIR](#).

Rec. 3: A model for FAIR Data Objects

Implementing FAIR requires a model for FAIR Data Objects which by definition have a PID linked to different types of essential metadata, including provenance and licencing. The use of community standards and sharing of code is also fundamental for interoperability and reuse.

- Universal use of appropriate PIDs needs to be facilitated and implemented.
Stakeholders: Data services; Institutions; Publishers; Funders.
- Educational programmes and tools are needed to raise awareness, understanding and use of relevant standards and routine capture of metadata during the research process.
Stakeholders: Data stewards; Institutions; Data services.
- Systems must be put in place for automatic checks on the existence and accessibility of PIDs, metadata, a licence or waiver, and code, and to test the validity of the links between them.
Stakeholders: Data services; Standards bodies.



DATA

The core bits

At its most basic level, data is a bitstream or binary sequence. For data to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and code. These layers of meaning enrich the data and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Data should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Data should be represented in common and ideally open file formats. This enables others to reuse the data as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

METADATA

Contextual documentation

In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest reuse, data should be accompanied by a 'plurality of relevant attributes' and a clear and accessible data usage license.

Figure 1: A model for FAIR Data Objects, noting the elements that need to be in place for data to be Findable, Accessible, Interoperable and Reusable.

Step 2: Develop and support a sustainable FAIR data ecosystem

Rec. 4: Components of a FAIR data ecosystem

The realisation of FAIR data relies on, at minimum, the following essential components: policies, DMPs, identifiers, standards and repositories. There need to be registries cataloguing each component of the ecosystem and automated workflows between them.

- Registries need to be developed and implemented for all of the FAIR components and in such a way that they know of each other's existence and interact. Work should begin by enhancing existing registries for policies, standards and repositories to make these comprehensive, and initiate registries for DMPs and identifiers.

Stakeholders: Data services; Standards bodies; Global coordination fora.

- By default, the FAIR ecosystem as a whole and individual components should work for humans and for machines. Policies and DMPs should be machine-readable and actionable.

Stakeholders: Data services; Global coordination fora; Policymakers.

- The infrastructure components that are essential in specific contexts and fields, or for particular parts of research activity, should be clearly defined.

Stakeholders: Research communities; Data stewards; Global coordination fora.

- Testbeds need to be used to continually evaluate, evolve, and innovate the ecosystem.

Stakeholders: Data services; Data stewards.

Related recommendations: [Rec. 5: Sustainable funding for FAIR components](#); [Rec. 25: Facilitate automated processing](#).

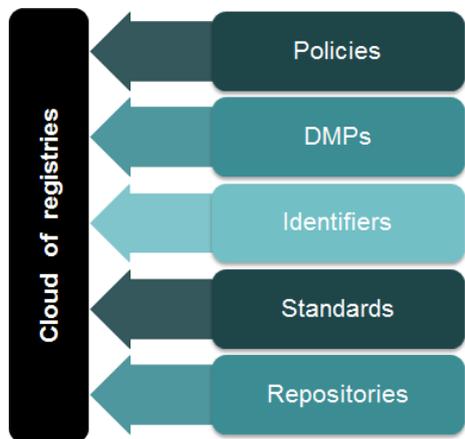


Figure 2: The components of a FAIR data ecosystem

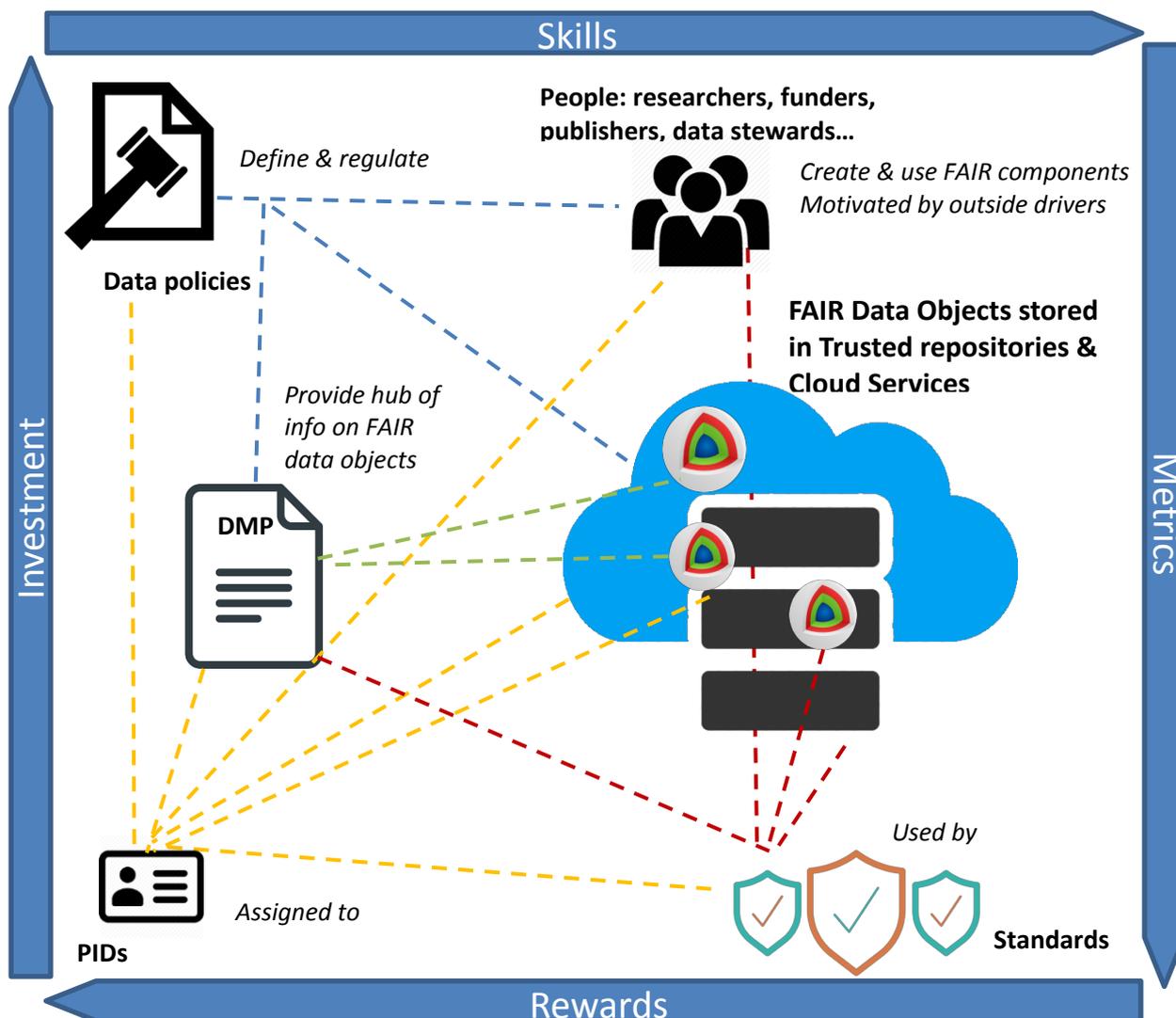


Figure 3: The interactions between components in the FAIR data ecosystem. Registries need to sit behind each component to support automated workflows across them.

Rec. 5: Sustainable funding for FAIR components

The components of the FAIR ecosystem need to be maintained at a professional service level with sustainable funding.

- Criteria for service acceptance and operation quality, including certification standards, need to be derived and applied with the aim to foster a systematic and systemic approach.
Stakeholders: Research communities; Global coordination fora; Funders.
- Regular evaluation of the relevance and quality of all services needed to support FAIR should be performed.
Stakeholders: Research communities; Data stewards.
- Sustainable funding and business models need to be developed for the provision of each of these components.
Stakeholders: Data services; Funders.

Related recommendations: [Rec. 33: Sustainable business models](#); [Rec. 11: Develop metrics to assess and certify data services](#).

Rec. 6: Strategic and evidence-based funding

Funders of research data services should consolidate and build on existing investments in infrastructure and tools, where they demonstrate impact and community adoption. Funding should be tied to certification schemes as they develop for each of the FAIR ecosystem components.

- Funding decisions for new and existing services should be tied to evidence, metrics and certification schemes validating service delivery.
Stakeholders: Funders; Institutions; Research communities.
- Effective guidance and procedures need to be established and implemented for retiring services that are no longer required (ref. [Principles for Open Scholarly infrastructures](#)).
Stakeholders: Data services; Data stewards.

Related recommendations: [Rec. 23: Incentivise services to support FAIR data](#); [Rec. 34: Leverage existing data services for EOSC](#).

Step 3: Ensure FAIR data and certified services to support FAIR

Rec. 7: Disciplinary interoperability frameworks

Research communities must be supported to develop and maintain their disciplinary interoperability frameworks. These incorporate principles and practices for data management and sharing, community agreements, data formats, metadata standards, tools and data infrastructure.

- Enabling mechanisms must be funded and implemented to support research communities to develop and maintain their disciplinary interoperability frameworks.
Stakeholders: Funders; Standards bodies; Data services; Global coordination fora.

- Disciplines and interdisciplinary research programmes should be encouraged to engage with international collaboration mechanisms to develop interoperability frameworks.
Stakeholders: Funders; Policymakers; Institutions; Data stewards; Global coordination fora.
- Mechanisms that promote the exchange of good practices and lessons learned within and across disciplines should be facilitated.
Stakeholders: Data services; Research communities; Global coordination fora.

Related recommendations: [Rec. 8: Cross-disciplinary FAIRness](#); [Rec. 16: Broad application of FAIR](#).

Rec. 8: Cross-disciplinary FAIRness

Interoperability frameworks should be articulated in common ways and adopt global standards where relevant to enable interdisciplinary research. Common standards, intelligent crosswalks, brokering mechanisms and machine-learning should all be explored to break down silos.

- Efforts should be made to identify information and practices that apply across research communities and articulate these in common standards that provide a baseline for FAIR.
Stakeholders: Standards bodies; Research communities.
- Case studies for cross-disciplinary data sharing and reuse should be collected. Based on these case studies, mechanisms that facilitate the development of frameworks for interoperability and reuse should be developed.
Stakeholders: Global coordination fora; Data stewards.
- The components of the FAIR ecosystem should adhere to common standards to support disciplinary frameworks and to promote interoperability and reuse of data across disciplines
Stakeholders: Data services; Research communities; Global coordination fora.

Related recommendations: [Rec. 7: Disciplinary interoperability frameworks](#).

Rec. 9: Develop robust FAIR data metrics

A set of metrics for FAIR Data Objects should be developed and implemented, starting from the basic common core of descriptive metadata, PIDs and access. The design of these metrics needs to be mindful of unintended consequences, and they should be regularly reviewed and updated.

- A core set of metrics for FAIR Data Objects should be defined to apply globally across research domains. More specific metrics should be defined at the community level to reflect the needs and practices of different domains and what it means to be FAIR for that type of research.
Stakeholders: Global coordination fora; Research communities.
- The European Commission should support a project to coordinate the activities of various groups defining FAIR metrics and ensure these are created in a standardised way to enable future monitoring.
Stakeholders: Funders.

- The process of developing, approving and implementing FAIR metrics should follow a consultative methodology, including scenario planning, to minimise to the greatest extent possible any unintended consequences and counter-productive gaming that may result. Metrics need to be regularly reviewed and updated to ensure they remain fit-for-purpose.

Stakeholders: Global coordination fora; Publishers; Data services.

Related recommendations: [Rec. 11: Develop metrics to assess and certify data services.](#)

Rec. 10: Trusted Digital Repositories

Repositories need to be encouraged and supported to achieve CoreTrustSeal certification. The development of rival repository accreditation schemes, based solely on the FAIR principles, should be discouraged.

- A programme of activity is required to incentivise and assist existing domain repositories, institutional services and other valued community resources to achieve CoreTrustSeal certification.

Stakeholders: Funders; Data services; Standards bodies.

- A transition period is needed to allow existing repositories without certifications to go through the steps needed to achieve trustworthy digital repository status. Concerted support is necessary to assist existing repositories in achieving certification.

Stakeholders: Data services; Institutions; Data stewards.

- At an appropriate point, the language of the CoreTrustSeal requirements should be reviewed and adapted to reference the FAIR data principles more explicitly (e.g. in sections on levels of curation, discoverability, accessibility, standards and reuse).

Stakeholders: Global coordination fora; Data services; Institutions.

- Repositories may need to adapt their services to enable and facilitate machine processing and to expose their holdings via standardised protocols.

Stakeholders: Data services; Institutions.

- CoreTrustSeal should also be supported to achieve scalability to meet the needs of repository certification in the FAIR context.

Stakeholders: Funders, Standards bodies.

- Mechanisms need to be developed to ensure that the repository ecosystem as a whole is fit for purpose, not just assessed on a per repository basis.

Stakeholders: Global coordination fora; Research communities.

Related recommendations: [Rec. 11: Develop metrics to assess and certify data services](#); [Rec. 18: Deposit in Trusted Digital Repositories.](#)

Rec. 11: Develop metrics to assess and certify data services

Certification schemes are needed to assess all components of the FAIR data ecosystem. Like CoreTrustSeal, these should address aspects of service management and sustainability, rather than being based solely on FAIR principles which are primarily articulated for data and objects.

- Building on the model of CoreTrustSeal, **new** certification schemes should be developed and refined by the community to assess and certify **other** core components needed in the FAIR data ecosystem, such as identifier services, standards and vocabularies.
Stakeholders: Global coordination fora; Data services; Standards bodies.
- Formal registries of certified components are needed: these must be maintained primarily by the certifying organisation, but should also be communicated in community discovery registries such as Re3data and FAIRsharing.
Stakeholders: Data services.
- Steps need to be taken to ensure that the organisations overseeing certification schemes are independent, trusted, sustainable and scalable.
Stakeholders: Funders; Research communities.

Related recommendations: [Rec. 10: Trusted Digital Repositories](#); [Rec. 9: Develop robust FAIR data metrics](#).

Step 4: Embed a culture of FAIR in research practice

Rec. 12: Data Management via DMPs

Any research project should include data management as a core element necessary for the delivery of its scientific objectives, addressing this in a Data Management Plan. The DMP should be regularly updated to provide a hub of information on the FAIR data objects.

- Research communities should be required and supported to consider data management and sharing as part of all research activities.
Stakeholders: Funders; Institutions; Data stewards; Publishers; Research communities.
- Data Management Plans should be living documents that are implemented throughout the project. A lightweight data management and curation statement should be assessed at project proposal stage, including information on costs and the track record in FAIR. A sufficiently detailed DMP should be developed at project inception. Project end reports should include reporting against the DMP.
Stakeholders: Funders; Institutions; Data stewards; Research communities.
- Research institutions and research projects need to take data management seriously and provide sufficient resources to implement the actions required in DMPs.
Stakeholders: Institutions; Data stewards; Research communities.
- Research communities should be inspired and empowered to provide input to the disciplinary aspects of DMPs and thereby to agree model approaches, exemplars and rubrics that help to embed FAIR data practices in different settings.
Stakeholders: Data services; data stewards; Research communities.

Related recommendations: [Rec. 21: Use information held in DMPs](#); [Rec. 32: Costing data management](#).

Rec. 13: Professionalise data science and data stewardship roles

Steps need to be taken to develop two cohorts of professionals to support FAIR data: data scientists embedded in those research projects which need them, and data stewards who will ensure the management and curation of FAIR data.

- Formal career pathways must be implemented to demonstrate the value of these roles and retain such professionalised roles in research teams.
Stakeholders: Institutions; Global coordination fora.
- Key data roles need to be recognised and rewarded, in particular, the data scientists who will assist research design and data analysis, visualisation and modelling; and data stewards who will inform the process of data curation and take responsibility for data management.
Stakeholders: Funders; Institutions; Publishers; Research communities.
- Professional bodies for these roles should be created and promoted. Accreditation should be developed for training and qualifications for these roles.
Stakeholders: Institutions; Data services; Research communities.

Related recommendations: [Rec. 28: Curriculum frameworks and training](#); [Rec. 14: Recognise and reward FAIR data and data stewardship](#).

Rec. 14: Recognise and reward FAIR data and data stewardship

FAIR data should be recognised as a core research output and included in the assessment of research contributions and career progression. The provision of infrastructure and services that enable FAIR data must also be recognised and rewarded accordingly.

- Policy guidelines should recognise the diversity of research contributions (including publications, datasets, online resources, teaching materials) at the level of biography and in templates for researchers' applications and activity reports.
Stakeholders: Funders; Publishers; Institutions.
- Credit should be given for all roles supporting FAIR data, including data analysis, annotation, management, curation and participation in the definition of disciplinary interoperability frameworks.
Stakeholders: Funders; Publishers; Institutions.
- Evidence of past practice in support of FAIR data should be included in assessments of research contribution. Such evidence should be required in grant proposals (for both research and infrastructure investments), for career advancement, for publication and conference contributions, and other evaluation schemes.
Stakeholders: Funders; Institutions; Publishers; Research communities.
- The contributions of organisations and collaborations to the development of certified and trusted infrastructures that support FAIR data should be recognised, rewarded and appropriately incentivised.
Stakeholders: Funders; Institutions.

Related recommendations: [Rec. 13: Professionalise data science and data stewardship roles](#).

FAIR data policy

In order to implement data policy effectively, we need a clear definition and understanding of FAIR. Related concepts such as Open Data which are already prevalent in policy need to be clarified in the context of FAIR. Policy should be harmonised to ease implementation, and the FAIR principles should be applied to a broad range of research objects. In addition to [Rec. 1: Definitions of FAIR](#) and [Rec. 2: Mandates and boundaries for Open](#), the following interventions are needed on data policy.

Rec. 15: Policy harmonisation

Efforts should be made to align and consolidate FAIR data policy, reducing divergence, inconsistencies and contradictions.

- Concerted work is needed to update policies to incorporate and align with the FAIR principles to ensure that policy properly supports the FAIR data Action Plan.
Stakeholders: Policymakers
- A funders' forum at a European and global level should do concrete work to align policies, DMP requirements and principles governing recognition and rewards.
Stakeholders: Funders.
- Information on practice in relation to exceptions should be captured and fed into a body of knowledge which can inform future policy guidance and practice.
Stakeholders: Policymakers; Global coordination fora.
- Policies should be versioned, indexed and semantically annotated in a policy registry.
Stakeholders: Policymakers; Data services; Global coordination fora.

Rec. 16: Broad application of FAIR

FAIR should be applied broadly to all objects (including metadata, identifiers, software and DMPs) that are essential to the practice of research, and should inform metrics relating directly to these objects.

- Policies must assert that the FAIR principles should be applied to research data, to metadata, to code, to DMPs and to other relevant digital objects.
Stakeholders: Policymakers.
- The FAIR data principles and this Action Plan must be tailored for specific contexts and the precise application nuanced, while respecting the objective of maximising data accessibility and reuse.
Stakeholders: Research communities; Data services; Policymakers.
- Guidelines for the implementation of FAIR in relation to research data, to metadata, to code, DMPs and other relevant digital objects should be developed and followed.
Stakeholders: Data services; Data stewards; Research communities; Funders.
- Examples and case studies of implementation should be collated so that other organisations can learn from good practice.
Stakeholders: Global coordination fora; Research communities.

Related recommendations: [Rec. 7: Disciplinary interoperability frameworks](#).

FAIR data culture

The primary actions needed to change research culture to embed FAIR practices are to support communities to develop interoperability frameworks (Rec. 7) and to specify these in ways that facilitate interdisciplinary research and prevent the formation of data silos (Rec. 8). All research projects should regard data management as a core component and address this in a Data Management Plan (Rec. 12). To facilitate the culture change needed, stakeholders that fund, publish, assess or in other ways legitimise research output, need to recognise and reward FAIR practices (Rec. 14).

Complementing these primary recommendations, a number of additional actions are suggested. Appropriate selection of research data of long-term value is critical to apply the Principles proportionally and ensure reusable materials are deposited in Trusted Digital Repositories. The FAIR principles are premised on access to and reuse of data, so this should be incentivised, and support offered to make legacy data FAIR where necessary. Since every research project will be creating a Data Management Plan, the information held in these should be reused to drive data exchange across the FAIR ecosystem.

Rec. 17: Selection and prioritisation of FAIR Data Objects

Research communities and data stewards should better define which FAIR data objects are likely to have long-term value and implement processes to assist the appraisal and selection of outputs that will be retained in the long term and made FAIR.

- Research communities should be encouraged and funded to make concerted efforts to improve guidance and processes on what to keep and make FAIR and what not to keep.
Stakeholders: Policymakers; Funders; Data services; Global coordination fora.
- The appraisal and selection of research outputs that are likely to have future research value and significance should reference current and past activities and emergent priorities.
Stakeholders: Research communities; Data stewards; Data services.
- When data are to be deleted as part of selection and prioritisation efforts, metadata about the data and about the deletion decision should be kept.
Stakeholders: Research communities; Data stewards; Data services.

Rec. 18: Deposit in Trusted Digital Repositories

Research data should be made available by means of Trusted Digital Repositories, and where possible in those with a mission and expertise to support a specific discipline or interdisciplinary research community.

- Policy should require data deposit in certified repositories and specify support mechanisms (e.g. incentives, funding of deposit fees, and training) to enable compliance.
Stakeholders: Policymakers; Funders; Publishers.
- Mechanisms need to be established to support research communities to determine the optimal data repositories and services for a given discipline or data type.
Stakeholders: Data services; Institutions; Data stewards.
- Concrete steps need to be taken to ensure the development of domain repositories and data services for interdisciplinary research communities so the needs of all researchers are covered.
Stakeholders: Data services; Funders; Institutions.

- Advocacy via scholarly societies, scientific unions and domain conferences is required so researchers in each field are aware of the relevant disciplinary repositories.

Stakeholders: Data services.

Related recommendations: [Rec. 10: Trusted Digital Repositories](#).

Rec. 19: Encourage and incentivise data reuse

Funders should incentivise data reuse by promoting this in funding calls and requiring research communities to seek and build on existing data wherever possible.

- Researchers should be required to demonstrate in DMPs that existing FAIR data resources have been consulted and used where possible before creating new data.

Stakeholders: Policymakers; Research communities.

- Appropriate levels of funding should be dedicated to reusing existing FAIR data by schemes that incentivise this.

Stakeholders: Funders; Institutions.

Rec. 20: Support legacy data to be made FAIR

There are large amounts of legacy data that is not FAIR but would have considerable value if it were. Mechanisms should be explored to include some legacy data in the FAIR ecosystem where required.

- Research communities and data owners should explore legacy data to identify indispensable collections with significant reuse potential that warrant effort to make them FAIR.

Stakeholders: Research communities; Institutions; Data services.

- Funding should be provided to adapt legacy datasets that have been identified as particularly crucial in a given discipline.

Stakeholders: Funders; Institutions; Research communities.

Rec. 21: Use information held in Data Management Plans

DMPs hold valuable information on the data and related outputs, which should be structured in a way to enable reuse. Investment should be made in DMP tools that adopt common standards to enable information exchange across the FAIR data ecosystem.

- DMPs should be explicitly referenced in systems containing information about research projects and their outputs (CRIS). Relevant standards and metadata profiles, should consider adaptations to include DMPs as a specific project output entity (rather than inclusion in the general category of research products). The same should apply to FAIR Data Objects.

Stakeholders: Standards bodies; Global coordination fora; Data services.

- A DMP standard should be developed that is extensible (e.g. like Dublin Core) by discipline (e.g. Darwin Core) or by the characteristics of the data (e.g. scale, sensitivity), or the data type (specific characteristics and requirements of the encoding).

Stakeholders: Standards bodies; Global coordination fora; Data services.

- Work is necessary to make DMPs machine readable and actionable. This includes the development of concepts and tools to support the creation of useful and usable data management plans tied to the actual research workflows.

Stakeholders: Funders; Data services; Data stewards.

- DMPs themselves should conform to FAIR principles and be Open where possible.
Stakeholders: Data services; Research communities; Policymakers.
- Information gathered from the process of implementing and evaluating DMPs relating to conformity, challenges and good practices should be used to improve practice.
Stakeholders: Data services; Funders; Research communities; Global coordination fora.

Related recommendations: [Rec. 4: Components of a FAIR data ecosystem](#); [Rec. 25: Facilitate automated processing](#).

Technology for FAIR

In order to support the implementation of the FAIR principles at a technical level, it is necessary to define the core elements of FAIR Data Objects ([Rec. 3](#)) and develop a FAIR data ecosystem comprising the necessary technical services to create, manage and share these objects in a FAIR way ([Rec. 4](#)). Some components of the ecosystem such as data repositories are already well advanced, with a wide-range of domain repositories available and existing mechanisms being adopted to certify the trustworthiness of these services ([Rec. 10](#)). For other components of the ecosystem, the metrics to assess and endorse services still need to be developed ([Rec. 11](#)).

In addition to these primary recommendations, it is also critical that the components being developed meet research needs and that services are incentivised to support FAIR data. To make the ecosystem interoperable and suitable for both human and machine access, we also need to support semantic technologies and facilitate automated processing.

Rec. 22: Develop FAIR components to meet research needs

While there is much existing infrastructure to build on, the further development and extension of FAIR components is required. These tools and services should fulfil the needs of data producers and users, and be easy to adopt.

- The development of FAIR compliant components needs to involve scientific communities, technical experts and other stakeholders. They should be provided with a forum for the exchange of views.
Stakeholders: Data services; Research communities; Global coordination fora.
- Engagement of the necessary stakeholders and experts needs to be facilitated with appropriate funding, support, incentives and training.
Stakeholders: Funders; Institutions.
- FAIR components will need regular iteration cycles and evaluation processes to ensure that they are fit for purpose and meet community needs.
Stakeholders: Data services; Research communities.

Related recommendations: [Rec. 7: Disciplinary interoperability frameworks](#); [Rec. 8: Cross-disciplinary FAIRness](#).

Rec. 23: Incentivise services to support FAIR data

Research facilities, in particular those of the ESFRI and national Roadmaps, should be incentivised to provide FAIR data by including it as a criteria in the initial and continuous evaluation process. Strategic research investments should consider service sustainability.

- The metrics and criteria by which research infrastructure are assessed should reference and build on the FAIR principles, incorporating language and concepts as appropriate, in order to align policy with implementation and to avoid confusion and dispersion of effort.

Stakeholders: Funders, Data services.

- Investment in new tools, services and components of the FAIR data ecosystem must be made strategically in order to leverage existing investments and ensure services are sustainable.

Stakeholders: Funders; Institutions.

Related recommendations: [Rec. 5: Sustainable funding for FAIR components](#); [Rec. 10: Trusted Digital Repositories](#); [Rec. 11: Develop metrics to assess and certify data services](#).

Rec 24: Support semantic technologies

Semantic technologies are essential for interoperability and need to be developed, expanded and applied both within and across disciplines.

- Programs need to be funded to make semantic interoperability more practical, including the further development of metadata standards, vocabularies and ontologies, along with appropriate validation infrastructure.

Stakeholders: Funders; Standards bodies; Global coordination fora.

- To achieve interoperability between repositories and registries, common protocols should be developed that are independent of the data organisation and structure of various services.

Stakeholders: Data services; Standards bodies.

Related recommendations: [Rec. 4: Components of a FAIR data ecosystem](#); [Rec. 8: Cross-disciplinary FAIRness](#).

Rec. 25: Facilitate automated processing

Automated processing should be supported and facilitated by FAIR components. This means that machines should be able to interact with each other through the system, as well as with other components of the system, at multiple levels and across disciplines.

- Automated workflows between the various components of the FAIR data ecosystem should be developed by means of coordinated activities and testbeds.

Stakeholders: Data services; Standards bodies.

- Metadata standards should be adopted and used consistently in order to enable machines to discover, assess and utilise data at scale.

Stakeholders: Data services; Research communities.

- Structured discoverability and profile matching mechanisms need to be developed and tested to broker requests and mediate metadata, rights, usage licences and costs.

Stakeholders: Data services.

Related recommendations: [Rec. 4: Components of a FAIR data ecosystem](#); [Rec. 8: Cross-disciplinary FAIRness](#); [Rec. 21: Use information held in Data Management Plans](#).

Skills and roles for FAIR

Both data science and data stewardship skills are needed for FAIR. Data science skills need to be core to research skills development and will often be used by researchers. However, with increasing specialisation, research groups may need to incorporate data scientists who assist with experimental design, statistics, data analysis, visualisation or modelling, and with machine learning in the case of particularly complex and large datasets. Data stewards who manage data, ensure that it is FAIR and prepare it for long term curation are also essential. Skills transfer schemes will be essential to ensure that professionals with sufficient subject knowledge are available for data science and data stewardship roles. Establishing curriculum frameworks and training programmes will help to establish the roles and achieve the primary objective, which is to professionalise data science and stewardship roles ([Rec. 13](#)).

Rec. 26: Data science and stewardship skills

Data skills of various types, as well as data management, data science and data stewardship competencies, need to be developed and embedded at all stages and with all participants in the research endeavour.

- Data skills, including an appropriate foundational level of in data science and data stewardship, should be included in undergraduate and postgraduate training across disciplines, and in the provision of continuing professional development (CPD) credits for researchers.
Stakeholders: Institutions; Data services; Research communities.
- More in-depth data science and data stewardship skills should be embedded in Master's degree courses for Information Professionals, so future generations of librarians, archivists and information systems staff are equipped to deal with the increasing complexity of research outputs.
Stakeholders: Institutions; Data services.

Rec. 27: Skills transfer schemes and brokering roles

Skills transfer schemes should be supported to equip researchers from various domains with information management skills or vice versa. Such individuals will play an important role as intermediaries to broker relations between research communities and infrastructure services.

- Investigate and learn from existing programmes that have demonstrated success in sharing skills across research scientist and information professional roles
Stakeholders: Funders; Institutions; Research communities.
- Support programmes of activity that enable skills transfer across communities.
Stakeholders: Funders; Institutions; Data services.

Rec. 28: Curriculum frameworks and training

A concerted effort should be made to coordinate, systematise and accelerate the pedagogy and availability of training for data skills, data science and data stewardship.

- Curriculum frameworks should be made available and be easily adaptable and reusable.
Stakeholders: Institutions.
- Sharing and reuse of Open Educational Resources and reusable materials for data science and data stewardships programmes should be encouraged and facilitated.
Stakeholders: Institutions; Global coordination fora; Data services.

- Train-the-Trainer programmes for data science and data stewardship roles should be developed, implemented and supported, so they can scale.

Stakeholders: Institutions; Data services; Data stewards; Funders.

- A programme of certification and endorsement should be developed for organisations and programmes delivering Train-the-Trainer and/or data science and data stewardship training. As a first step, a lightweight peer-reviewed self-assessment would be a means of accelerating the development and implementation of quality training.

Stakeholders: Institutions; Global coordination fora; Standards bodies.

Related recommendation: [Rec. 13: Professionalise data science and data stewardship roles.](#)

FAIR metrics

In order to ‘Turn FAIR data into reality’, the concept and principles of FAIR need to become part of data policy requirements and research assessment frameworks. Underpinning the implementation of these policies and assessments are a robust set of metrics to validate that data are FAIR ([Rec. 9](#)) and that services are certified and support FAIR data ([Rec. 11](#)). These two primary recommendations are prerequisites for the implementation and monitoring of FAIR.

*The FAIR principles are articulated for data and related objects, and the development of metrics for FAIR Data Objects is already underway. Many elements of the FAIR data principles are relevant for services too, and should be incorporated into accreditation schemes. However, these schemes should also assess trustworthiness, sustainability and robust management practices. Foundational repository accreditation is provided by CoreTrustSeal, so there is no need for new FAIR-based certifications. New accreditation schemes **are** needed for other services that contribute to the FAIR ecosystem of components, such as identifier services, standards and vocabularies.*

Traditional metrics should also be enriched through next generation metrics and data citation, and the citation of FAIR Data Objects should be implemented in the scholarly literature for attribution and in research assessment frameworks for recognition and career advancement.

Rec. 29: Implement FAIR metrics

Agreed sets of metrics should be implemented and monitored to track changes in the FAIRness of datasets or data-related resources over time.

- Repositories should publish assessments of the FAIRness of datasets, where practical, based on community review and the judgement of data stewards. Methodologies for assessing FAIR data need to be piloted and developed into automated tools before they can be applied across the board by repositories.

Stakeholders: Data services; Institutions; Publishers.

- Metrics for the assessment of research contributions, organisations and projects should take the past FAIRness of datasets and other related outputs into account. This can include citation metrics, but appropriate alternatives should also be found for the research / researchers / research outputs being assessed.

Stakeholders: Funders; Institutions.

Related recommendations: [Rec. 9: Develop robust FAIR data metrics](#); [Rec. 11: Develop metrics to assess and certify data services](#); [Rec 14: Recognise and reward FAIR data and data stewardship.](#)

Rec. 30: Monitor FAIR

Funders should report annually on the outcomes of their investments in FAIR and track how the landscape matures. Specifically, how FAIR are the research objects that have been produced and to what extent are the funded infrastructures certified and supportive of FAIR data.

- Statistics should be published on the outcome of all investments to report on levels of FAIR data and certified services
Stakeholders: Funders; Institutions.
- The results of monitoring processes should be used to inform and iterate data policy.
Stakeholders: Policymakers; Funders; Institutions.

Rec. 31: Support data citation and next generation metrics

Systems providing citation metrics for FAIR Data Objects and other research outputs should be provided. In parallel, next generation metrics that reinforce and enrich citation-centric metrics for evaluation should be developed.

- Citation of data and other research outputs needs to be encouraged and supported, for example by including sections in publishing templates that prompt researchers to reference materials, and providing citation guidelines when data, code or other outputs are accessed.
Stakeholders: Publishers; Data services; Institutions.
- The Joint Data Citation Principles should be actively endorsed and implemented in the scholarly literature for attribution and in research assessment frameworks for recognition and career advancement.
Stakeholders: Publishers, Institutions, Funders.
- A broader range of metrics should be developed to recognise contributions beyond publications and citation. These should recognise and reward Open and FAIR data practices.
Stakeholders: Funders; Publishers; Institutions.

Related recommendation: [Rec. 14: Recognise and reward FAIR data and data stewardship](#); [Rec. 19: Encourage and incentivise data reuse](#).

Costs and investment in FAIR

Researchers can not be expected to make their data FAIR without appropriate tools, services and infrastructure. Policymakers should not introduce requirements without also investing in support to enable compliance. Moreover, the FAIR data ecosystem needs to be sustainably funded ([Rec. 5](#)), and that funding should be strategic and evidence-based ([Rec. 6](#)) to ensure the services are fit-for-purpose and meet community needs. These primary requirements should be supported by interventions to enable data management costs to be calculated and included in proposals, and sustainable business models to be explored by services. Existing services should also be used where appropriate to make the best use of investment and avoid reinventing wheels.

Rec. 32: Costing data management

Research funders should require data management costs to be considered and included in grant applications, where relevant. To support this, detailed guidelines and worked examples of eligible costs for FAIR data should be provided.

- Details on the costs of data management, curation and publication should be included in all DMP templates.

Stakeholders: Funders, Institutions, Data services.

- Guidelines should be provided for researchers and reviewers to raise awareness of eligible costs and reinforce the view that data management, long term curation and data publication should be included in project proposals.

Stakeholders: Funders; Institutions.

- Information from existing and completed projects should be used to retrospectively identify costs and develop examples and guidelines based on these.

Stakeholders: Funders; Institutions; Data services.

Related recommendations: [Rec. 12: Data management via DMPs.](#)

Rec. 33: Sustainable business models

Data repositories and other components of the FAIR data ecosystem should be supported to explore business models for sustainability, to articulate their value proposition, and to trial a range of charging models and income streams.

- Examples of different business models should be shared, and data services given time and support to trial approaches to test the most viable sustainability paths.

Stakeholders: Funders; Data services; Global coordination bodies.

Related recommendations: [Rec. 5: Sustainable funding for FAIR components;](#) [Rec. 32: Costing data management.](#)

Rec. 34: Leverage existing data services for EOSC

The Rules of Engagement for EOSC must be broadly-defined and open to enable all existing service providers to address the criteria and be part of the European network.

- The Rules of Engagement for EOSC must be consulted on widely, drawing in views from a broad range of stakeholder groups beyond the core European Research Infrastructures and E-Infrastructures to include research communities, institutions, publishers, commercial service providers and international perspectives.

Stakeholders: Data services; Research communities; Institutions; Publishers.

- The resulting Rules must be fit-for-purpose to enable all existing data services and capacities developed by different communities to be exploited for best return on investment. The Rules should be reviewed regularly to ensure they remain viable.

Stakeholders: Data services; Research communities; Policymakers.

Related recommendations: [Rec. 6: Strategic and evidence-based funding.](#)

How the FAIR Data Action Plan supports the EOSC

As noted in the European Commission's Staff Working Document providing an *Implementation Roadmap for the European Open Science Cloud*, the FAIR data Action Plan is intended to set out the actions needed to develop EOSC shared resources and define the operational guidance and methodologies for applying the FAIR principles with these shared resources. Some recommendations apply directly, for example [Recommendation 34: Leverage existing data services for EOSC](#) to ensure the Rules of Engagement are sufficiently broad and ratified by community consensus. Most of the recommendations in the FAIR Data Action Plan, however, are intentionally articulated more broadly to apply to member states and the international community, since research is global.

The framework proposed for FAIR Data Objects supported by a FAIR data ecosystem that addresses the cultural and technical developments needed should be used to guide the operation of the EOSC. The recommendations and actions propose the changes required on a policy, cultural and technical level to support FAIR and embed these practices across research communities. The implementation path pursued by the EOSC should be done in parallel with similar activities internationally, such as the NIH Data Commons, the African Open Science Platform and the Australian NeCTAR Research Cloud. Global coordination fora should be used to exchange experiences and ensure the services developed in Europe are interoperable internationally.