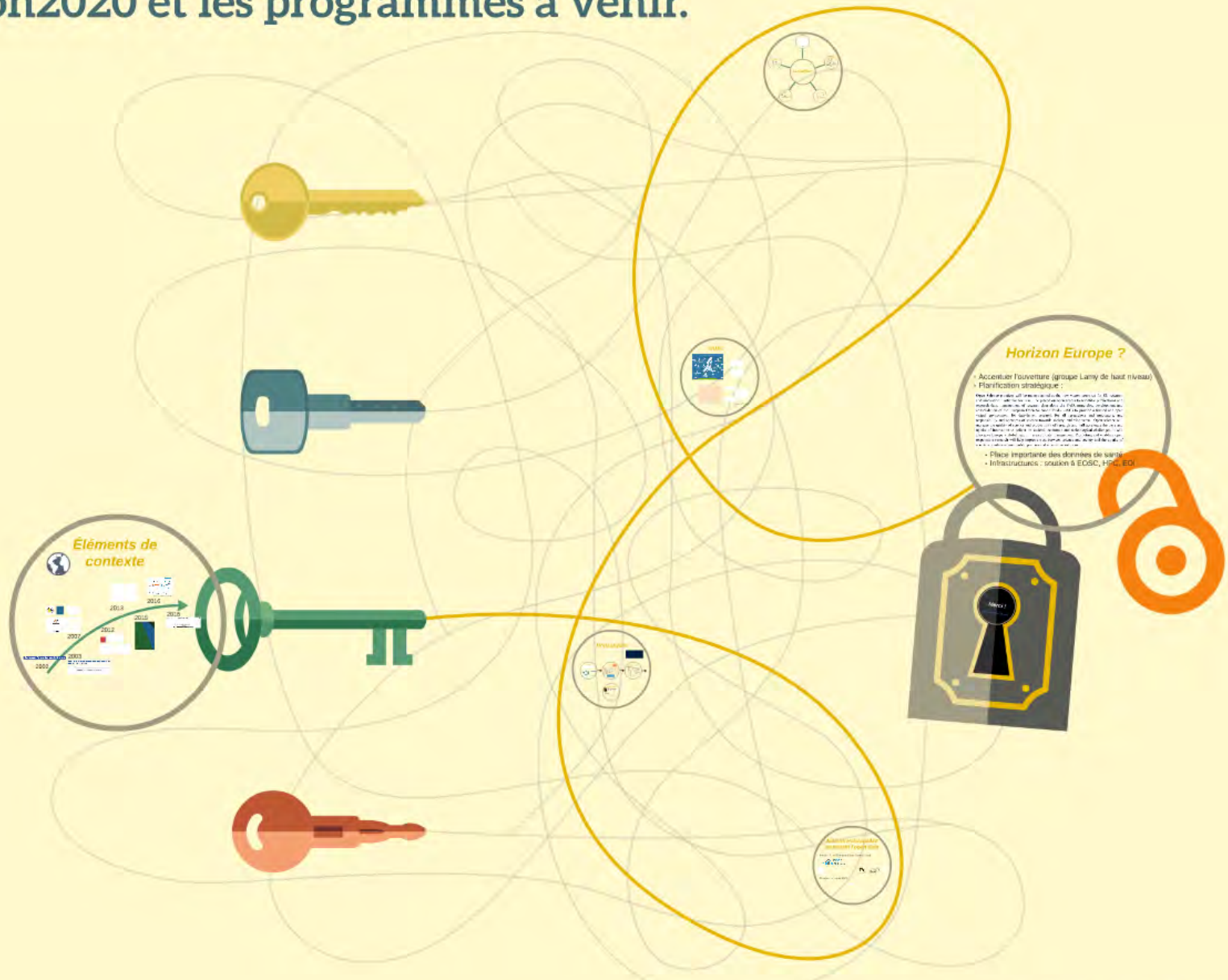
Enjeux européens autour de la question des données. Les DMPs, Horizon2020 et les programmes à venir.



Corinne Joffre, Université de Toulouse
dmp2019 : Formation DMP Grenoble - Toulouse, 27/06/2019

Éléments de contexte

2002

Budapest Open Access Initiative

2003

2007

2012

2013

2015

2016

2018

INSPIRE - Infrastructure for Spatial Information in Europe

COMMISSION DES COMMUNAUTÉS EUROPÉENNES

Bruxelles, le 14.2.2007
COM(2007) 56 final

COMMUNICATION DE LA COMMISSION AU PARLEMENT EUROPÉEN, AU CONSEIL ET AU COMITÉ ÉCONOMIQUE ET SOCIAL EUROPÉEN

sur l'information scientifique à l'ère numérique : accès, diffusion et préservation

{SEC(2007)181}

erc

European Research Council

Scientific Council

Established by the European Commission

ERC Scientific Council Guidelines for Open Access

17 December 2007

The ERC considers essential that primary data - which in the life sciences for example could comprise data such as nucleotide/protein sequences, macromolecular atomic coordinates and anonymized epidemiological data - are deposited to the relevant databases as soon as possible, preferably immediately after publication and in any case not later than 6 months after the date of publication.

Access to unprocessed data is needed not only for independent verification of results but, more importantly, for secure preservation and fresh analysis and utilisation of the data.

A number of freely accessible repositories and curated databases for publications and data already exist serving researchers in the EU. Over 400 research repositories are run by European research institutions and several fields of scientific research have their own international discipline-specific repositories. These include for example PubMed Central for peer-reviewed publications in the life sciences and medicine, the arXiv Internet preprint archive for physics and mathematics, the DDBJ/EMBL/GenBank nucleotide sequence database and the RSCB-PDB/MSD-EBI/PDBj protein structure database.

With few exceptions, the social sciences & humanities (SSH) do not yet have the benefit of public central repositories for their recent journal publications. The importance of open access to primary data, old manuscripts, collections and archives is even more acute for SSH. In the social sciences many primary or secondary data, such as social survey data and statistical data, exist in the public domain, but usually at national level. In the case of the humanities, open access to primary sources (such as archives, manuscripts and collections) is often hindered by private (or even public or nation-state) ownership which permits access either on a highly selective basis or not at all.

2012

COMMISSION EUROPÉENNE

Bruxelles, le 17.7.2012
COM(2012) 401 final

COMMUNICATION DE LA COMMISSION AU PARLEMENT EUROPÉEN, AU CONSEIL, AU COMITÉ ÉCONOMIQUE ET SOCIAL EUROPÉEN ET AU COMITÉ DES RÉGIONS

Pour un meilleur accès aux informations scientifiques: dynamiser les avantages des investissements publics dans le domaine de la recherche

COMMISSION EUROPÉENNE

Bruxelles, le 17.7.2012
C(2012) 4890 final

RECOMMANDATION DE LA COMMISSION

du 17.7.2012

relative à l'accès aux informations scientifiques et à leur conservation

{SWD(2012) 221 final}
{SWD(2012) 222 final}

COMMISSION EUROPÉENNE

Bruxelles, le 17.7.2012
COM(2012) 392 final

COMMUNICATION DE LA COMMISSION AU PARLEMENT EUROPÉEN, AU CONSEIL, AU COMITÉ ÉCONOMIQUE ET SOCIAL EUROPÉEN ET AU COMITÉ DES RÉGIONS

Un partenariat renforcé pour l'excellence et la croissance dans l'Espace européen de la recherche

(Texte présentant de l'intérêt pour l'EEE)

{SWD(2012) 211 final}
{SWD(2012) 212 final}

Éléments de contexte

2002

Budapest Open Access Initiative

2003

2007

2012

2013

2015

2016

2018

Horizon2020

2014

2016

Extension du pilote ORD à toutes les thématiques du programme :

- Les données de la recherche sont ouvertes par défaut
- Opt-out total ou partiel, dès la soumission du projet ou durant l'exécution du projet
- Les propositions doivent expliquer pourquoi elles refusent d'ouvrir leurs données
- Approche "As open as possible, as closed as necessary"
- Nouveau modèle de DMP introduisant le principe FAIR

Montage de projets

2. Impact
2.2 Measures to maximize impact
a) Dissemination and exploitation of results
As relevant, include information on how the participants will manage the research data generated and/or collected during the project, in particular addressing the following issues:
What types of data will the project generate/collect? What standards will be used? How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why. How will this data be curated and preserved? How will the costs for data curation and preservation be covered?

3. Implementation
3.1 Work plan
Inclure le DMP comme livrable avant M6.

Pendant le projet

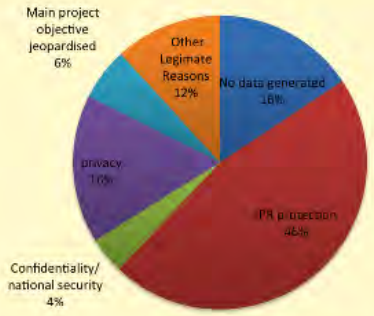
Prezi

2014

A novelty in Horizon 2020 is the Open Research Data Pilot which aims to improve and maximise access to and re-use of research data generated by projects. Participating projects will make their research data available on a voluntary basis, as specified in their Data Management Plans (DMPs, see below). They will also be required to make available the data needed to validate the results presented in scientific publications. Participating projects will receive dedicated support. In particular, any costs relating to the implementation of the pilot will be reimbursed and specific technical and professional support services will be provided.

Horizon2020, programme de travail 2014-2015 adopté le 10 décembre 2013

	Opt-out		Opt-in	
	Projets soumis	Projets financés	Projets soumis	Projets financés
2014-2015	22,26%	32,56%	3,12%	2,50%
2015-2016	17,87%	31,87%	17,95%	15,25%



Source: http://data.europa.eu/euodp/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020

Participating projects will receive dedicated support. In particular, any costs relating to the implementation of the pilot will be reimbursed and specific technical and professional support services will be provided.

Horizon2020, programme de travail 2014-2015 adopté le 10 décembre 2013

3 PRIORITÉS

EXCELLENCE SCIENTIFIQUE

- Conseil européen de la recherche (E.R.C.)
- Actions Marie Sklodowska-Curie
- Technologies futures et émergentes (FET)
- Infrastructures de recherche

PRIMAUTÉ INDUSTRIELLE

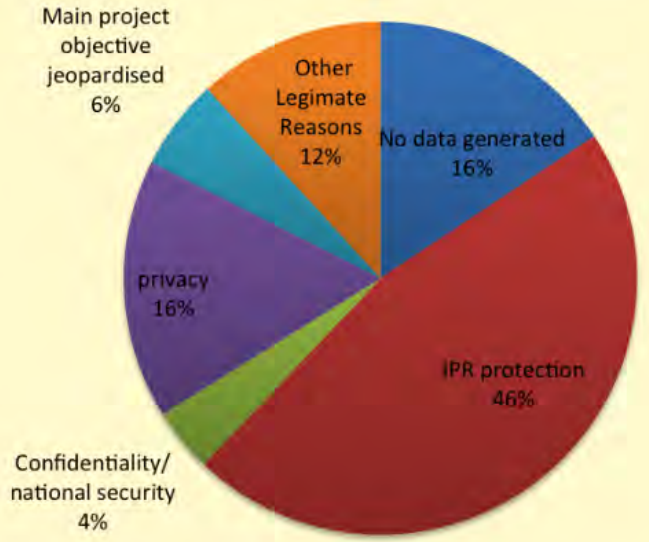
- TIC
- Technologies clés génériques (KET) :
  - microélectronique
  - photonique
  - nanotechnologies
  - matériaux avancés
  - systèmes de production
  - biotechnologies
- Espace
- Innovation dans les P.M.E.
- Accès au financement à risque

DÉFIS SOCIÉTAUX

- Santé, bien-être, vieillissement
- Sécurité alimentaire, bioéconomie...
- Energies sûres, propres, efficaces
- Transports intelligents, verts, intégrés
- Climat, environnement, matières premières
- Sociétés inclusives et novatrices et capables de réflexion
- Sociétés sûres

- Diffusion de l'excellence et élargissement de la participation
- Science pour et avec la société
- Institut Européen d'Innovation et Technologie (I.E.T.)
- Centre commun de recherche (Joint Research Center - J.R.C.)

	Opt-out		Opt-in	
	Projets soumis	Projets financés	Projets soumis	Projets financés
2014-2015	22,26%	32,56%	3,12%	2,50%
2015-2016	17,87%	31,87%	17,95%	15,25%



Source: http://data.europa.eu/euodp/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020

2016

Extension du pilote ORD à toutes les thématiques du programme :

- Les données de la recherche sont ouvertes par défaut
- Opt-out total ou partiel, dès la soumission du projet ou durant l'exécution du projet
- Les propositions doivent expliquer pourquoi elles refusent d'ouvrir leurs données
- Approche "As open as possible, as closed as necessary"
- Nouveau modèle de DMP introduisant le principe FAIR

Montage de projets

2. Impact

2.2 Measures to maximize impact

a) Dissemination and exploitation of results

As relevant, include information on how the participants will manage the research data generated and/or collected during the project, in particular addressing the following issues:

What types of data will the project generate/collect? What standards will be used? How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why. How will this data be curated and preserved? How will the costs for data curation and preservation be covered?

3. Implementation

3.1 Work plan

Inclure le DMP comme livrable avant M6.

Proposal ID	Acronym

5 - Call specific questions

Extended Open Research Data Pilot in Horizon 2020

If selected, applicants will by default participate in the Pilot on Open Research Data in Horizon 2020', which aims to improve and maximise access to and re-use of research data generated by actions.

However, participation in the Pilot is flexible in the sense that it does not mean that all research data needs to be open. After the action has started, participants will formulate a Data Management Plan (DMP), which should address the relevant aspects of making data FAIR – findable, accessible, interoperable and re-usable, including what data the project will generate, whether and how it will be made accessible for verification and re-use, and how it will be curated and preserved. Through this DMP projects can define certain datasets to remain closed according to the principle "as open as possible, as closed as necessary". A Data Management Plan does not have to be submitted at the proposal stage.

Furthermore, applicants also have the possibility to opt out of this Pilot completely at any stage (before or after the grant signature). In this case, applicants must indicate a reason for this choice (see options below).

Please note that participation in this Pilot does not constitute part of the evaluation process. Proposals will not be penalised for opting out.

We wish to opt out of the Pilot on Open Research Data in Horizon 2020.	○ Yes	◉ No

If opting out please indicate the reason(s) for not being able to participate in the Pilot:

- the project does not generate any data	☐
- to allow the protection of results (e.g. patenting)	☐
- incompatibility with the need for confidentiality linked to security	☐
- incompatibility with privacy/data protection	☐
- achievement of the project's main aim would be jeopardised	☐
- other legitimate reasons	☐

Please specify the reason:

Remaining characters 300

Montage de projets

2. Impact

2.2 Measures to maximize impact

a) Dissemination and exploitation of results

As relevant, include information on how the participants will manage the research data generated and/or collected during the project, in particular addressing the following issues:

What types of data will the project generate/collect? What standards will be used? How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why. How will this data be curated and preserved? How will the costs for data curation and preservation be covered?

3. Implementation

3.1 Work plan

Inclure le DMP comme livrable avant M6.

Pendant le projet

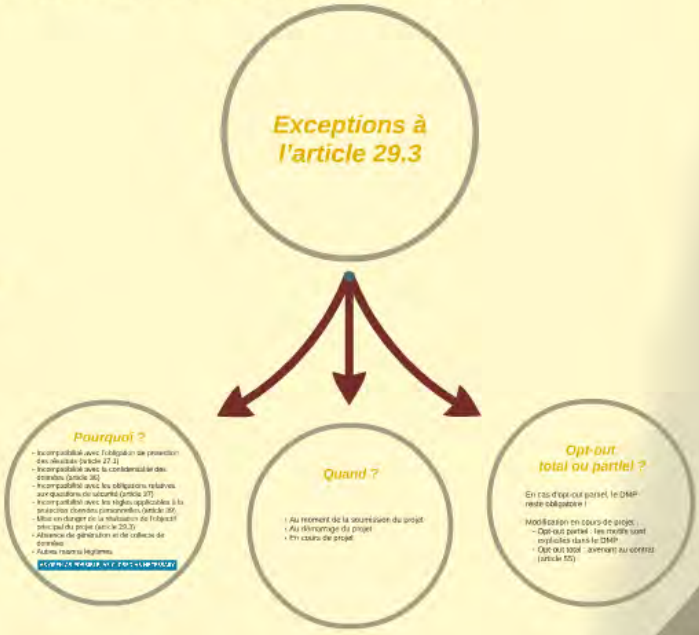
Obligation contractuelle (article 29.3) :

- déposer, le plus tôt possible, les données de la recherche nécessaires à la validation des résultats présentés dans les publications et l'ensemble des données visées dans le data management plan dans un répertoire de données de la recherche et présenter les instruments et outils nécessaires à la validation des résultats
- permettre aux tiers d'accéder, extraire/explorer (data mining), exploiter, reproduire et disséminer les données gratuitement

Deux types de données :

- Données et métadonnées nécessaires à la validation des publications revues par les pairs : toujours obligatoire
- Autres données et métadonnées, au choix du bénéficiaire : données et conditions spécifiées dans le DMP

Le DMP est un livrable du projet et devient une obligation contractuelle.



Pourquoi ?

- Incompatibilité avec l'obligation de protection des résultats (article 27.1)
- Incompatibilité avec la confidentialité des données (article 36)
- Incompatibilité avec les obligations relatives aux questions de sécurité (article 37)
- Incompatibilité avec les règles applicables à la protection données personnelles (article 39)
- Mise en danger de la réalisation de l'objectif principal du projet (article 29.3)
- Absence de génération et de collecte de données
- Autres raisons légitimes

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

Quand ?

- Au moment de la soumission du projet
- Au démarrage du projet
- En cours de projet

Opt-out total ou partiel ?

En cas d'opt-out partiel, le DMP reste obligatoire !

Modification en cours de projet :

- Opt-out partiel : les motifs sont explicités dans le DMP
- Opt-out total : avenant au contrat (article 55)

Horizon2020

2014

2016

Extension du pilote ORD à toutes les thématiques du programme :

- Les données de la recherche sont ouvertes par défaut
- Opt-out total ou partiel, dès la soumission du projet ou durant l'exécution du projet
- Les propositions doivent expliquer pourquoi elles refusent d'ouvrir leurs données
- Approche "As open as possible, as closed as necessary"
- Nouveau modèle de DMP introduisant le principe FAIR

Montage de projets

2. Impact
2.2 Measures to maximize impact
a) Dissemination and exploitation of results
As relevant, include information on how the participants will manage the research data generated and/or collected during the project, in particular addressing the following issues:
What types of data will the project generate/collect? What standards will be used? How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why. How will this data be curated and preserved? How will the costs for data curation and preservation be covered?

3. Implementation
3.1 Work plan
Inclure le DMP comme livrable avant M6.

Pendant le projet

Prezi

Actions Horizon2020 soutenant l'open data

Infrastructure : European Open Science Cloud

EOSCpilot
The European Open Science Cloud for Research Pilot Project

EOSC-hub

OpenAIRE

eInfra Central

RDA
RESEARCH DATA ALLIANCE

GÉANT

HELIX NEBULA
THESCIENCECLOUD

Pratique : principes FAIR

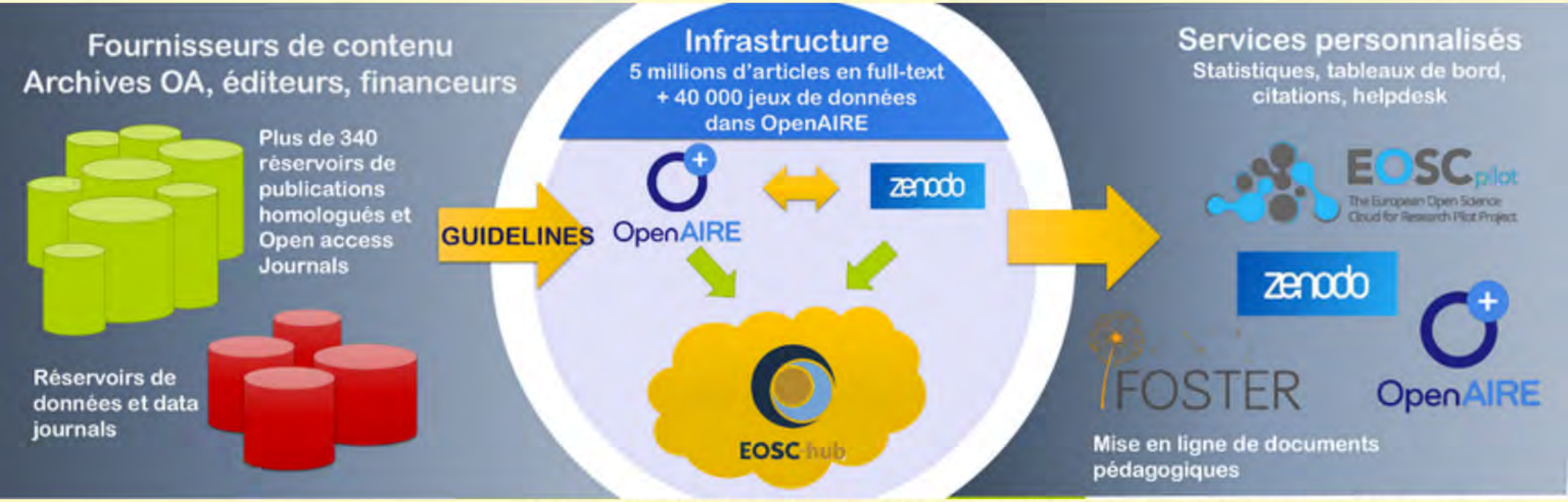
FAIRSFAIR
Fostering Fair Data Practices in Europe

OpenAIRE

GO FAIR

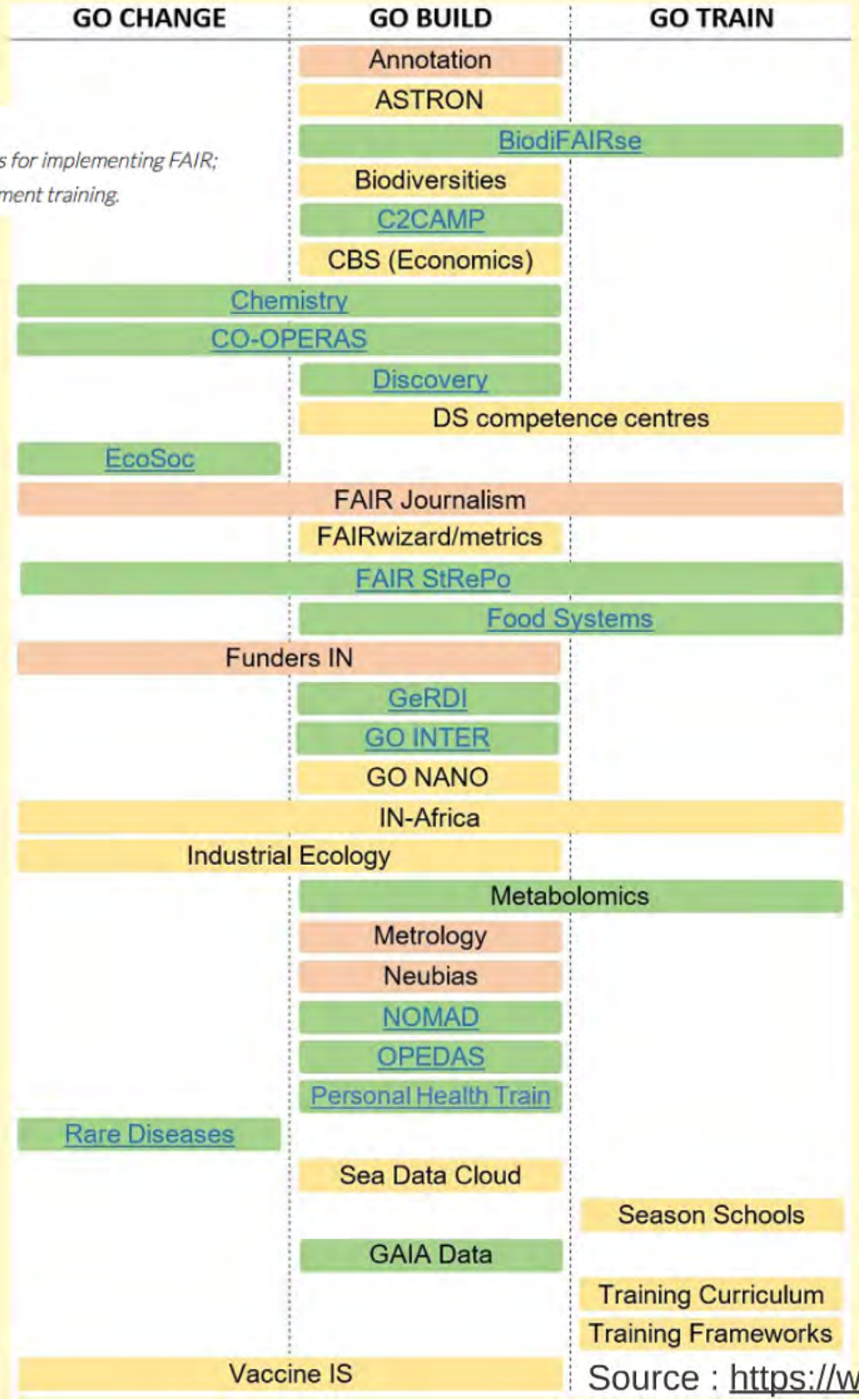
Source : https://www.fairsfair.eu/about-us

Collaboration avec EOSC et les clusters d'ESFRI (SSHOC, PANOSC, ENVRI FAIR, ESCAPE and EOSCLife)



Source : https://www.couperin.org/services-et-prospective/open-access/open-aire/item/1334-openaire-advance

***GO BUILD**, focus on FAIR technology;*
***GO CHANGE**, focus on priorities, policies and incentives for implementing FAIR;*
***GO TRAIN**, focus on FAIR awareness and skills development training.*



Source : https://www.go-fair.org/implementation-net

Actions Horizon2020 soutenant l'open data

Infrastructure : European Open Science Cloud

EOSCpilot
The European Open Science Cloud for Research Pilot Project

EOSC-hub

OpenAIRE

eInfra Central

RDA
RESEARCH DATA ALLIANCE

GÉANT

HELIX NEBULA
THESCIENCECLOUD

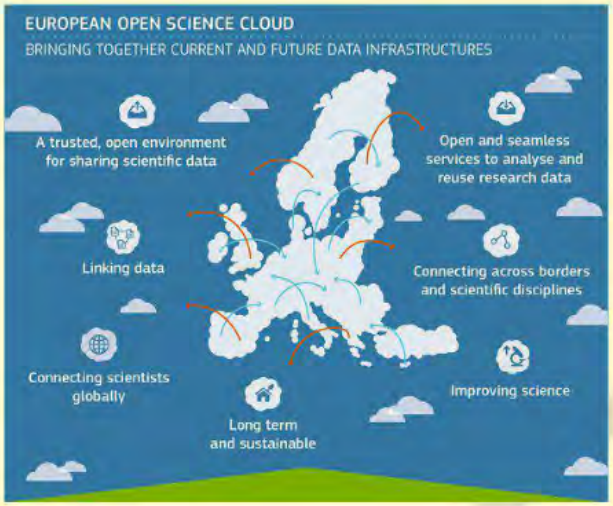
Pratique : principes FAIR

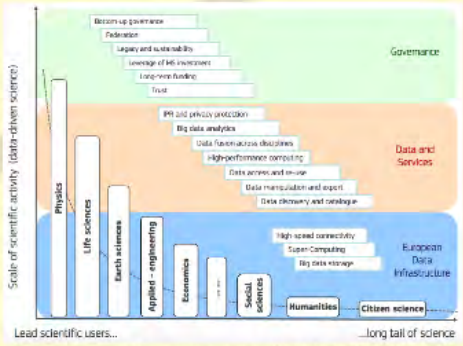
FAIRSFAIR
Fostering Fair Data Practices in Europe

OpenAIRE

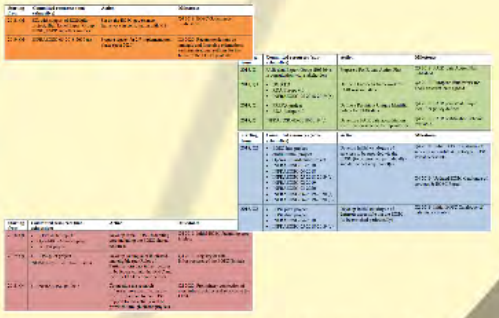
GO FAIR

EOSC





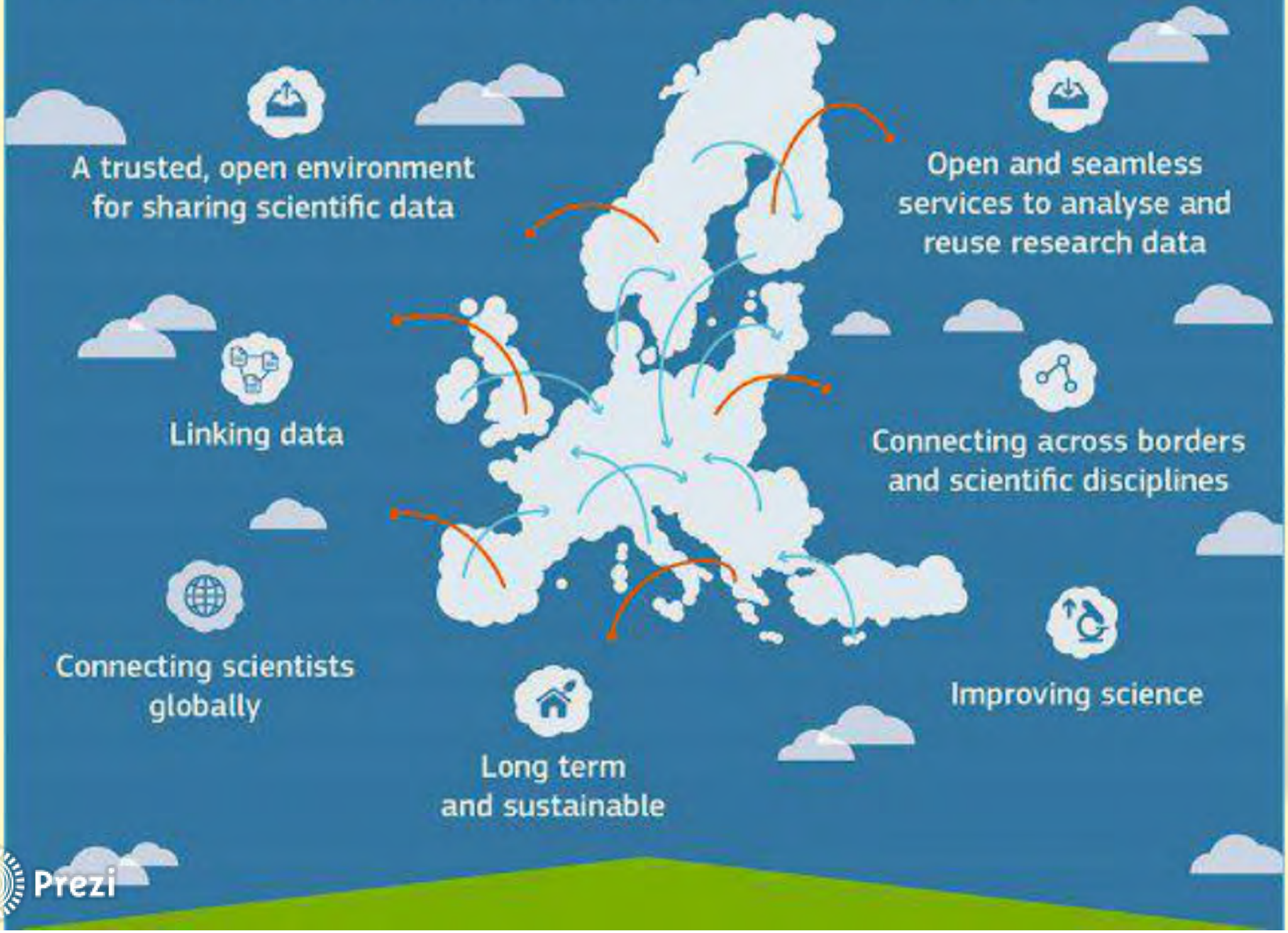
Workshop on governance and funding for the EOSC, 29 juin 2016



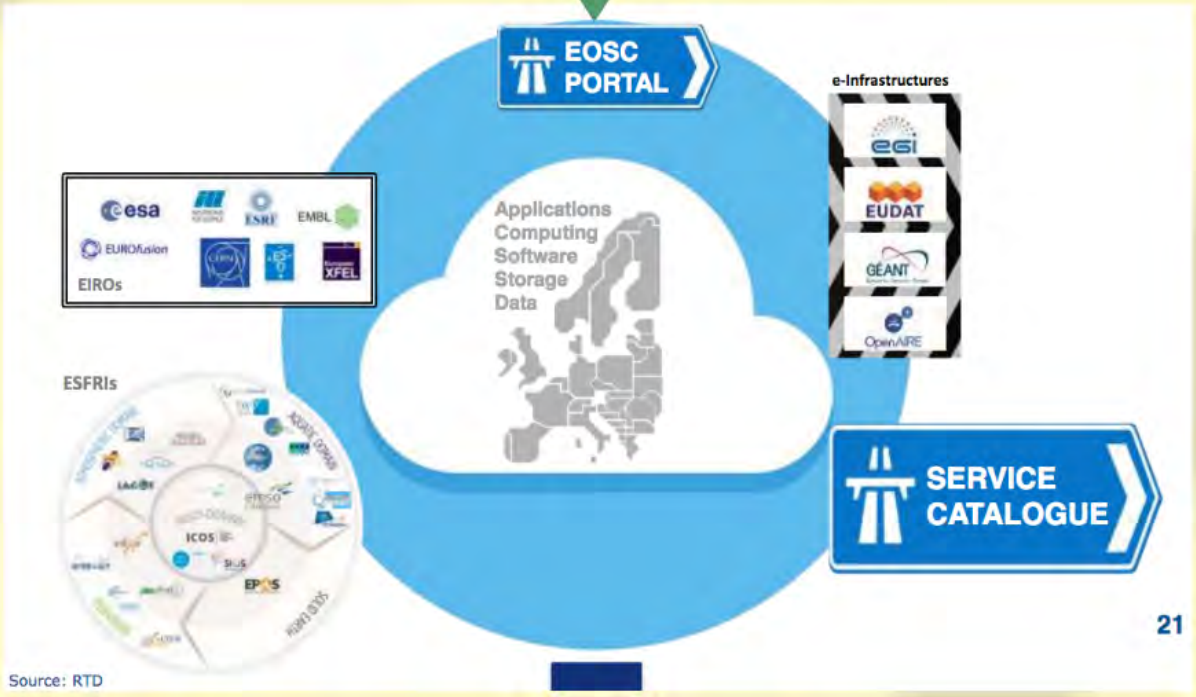
COMMISSION STAFF WORKING DOCUMENT Implementation Roadmap for the European Open Science Cloud, 3 mars 2017

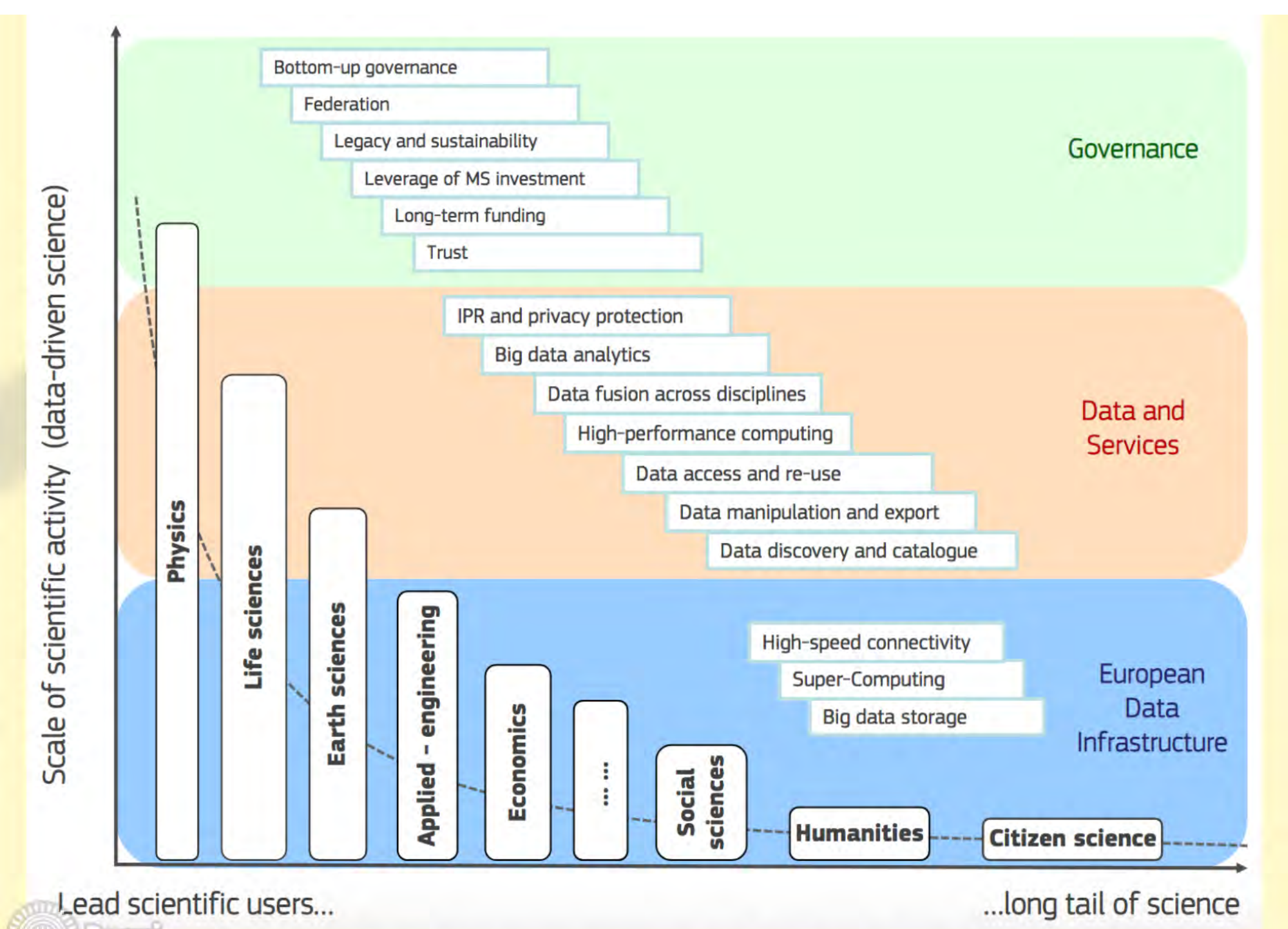
EUROPEAN OPEN SCIENCE CLOUD

BRINGING TOGETHER CURRENT AND FUTURE DATA INFRASTRUCTURES




Data
Computing
Storage
Applications
Software
Public Institutions
Universities
Libraries
EIROs
e-Infrastructures
ESFRIs
Commercial providers
LOCAL ACCESS ONLY
STOP
NO ENTRY
19
Source: RTD


EOSC PORTAL
EIROs
e-Infrastructures
Applications
Computing
Software
Storage
Data
ESFRIs
SERVICE CATALOGUE
21
Source: RTD


Bottom-up governance
Federation
Legacy and sustainability
Leverage of MS investment
Long-term funding
Trust
Governance
IPR and privacy protection
Big data analytics
Data fusion across disciplines
High-performance computing
Data access and re-use
Data manipulation and export
Data discovery and catalogue
Data and Services
High-speed connectivity
Super-Computing
Big data storage
European Data Infrastructure
Scale of scientific activity (data-driven science)
Physics
Life sciences
Earth sciences
Applied - engineering
Economics
... ...
Social sciences
Humanities
Citizen science
Lead scientific users...
...long tail of science

Starting from	Committed resources (non-exhaustive)	Action	Milestones
2018, Q1	EC, with support of EOSCpilot project, High Level Expert Group EOSC, OSPP and other sources	Set up the EOSC governance framework in consultation with MS	Q4 2018: EOSC Governance established
2019, Q1	INFRAEOSC-05-2018-2019 (a)	Prepare legacy for 2nd implementation phase (post 2020)	Q3 2020: Recommendations on strategic and financing orientations and organisational settings for the future of the EOSC, post 2020

Starting from	Committed resources (non-exhaustive)	Action	Milestones
2018, Q1	FAIR data Expert Group (E03464), in consultation with stakeholders	Prepare a FAIR data Action Plan	Q3 2018: FAIR data Action Plan published
2018, Q3	• DG RTD • RDA Europe 4.0 • INFRAEOSC-05-2018-2019 (b)	Define a European framework for FAIR research data	Q2 2019: European framework for FAIR research data agreed
2019, Q1	• FREYA project • RDA Europe 4.0	Define a Persistent Unique Identifier policy for FAIR data	Q4 2019: FAIR persistent unique identifier policy defined
2019, Q1	INFRAEOSC-05-2018-2019 (c)	Develop a FAIR data accreditation /certification scheme for repositories	Q4 2019: FAIR certification scheme available.

Starting from	Committed resources (non-exhaustive)	Action	Milestones
2018, Q2	• EOSC-hub project • eInfraCentral project • OpenAIRE-Advance project • INFRAEOSC-01-2018 • INFRAEOSC-04-2018 • INFRAEOSC-05-2018-2019 (b) • INFRAEOSC-02-2019 • INFRAEOSC-03-2020 • INFRAEOSC-06-2019-2020 (a) • INFRAEOSC-06-2019-2020 (b)	Develop **initial catalogue of services** to be provided via the EOSC (to be enriched periodically) and define delivery model(s)	Q4 2018: Initial EOSC Catalogue of services accessible & prototype EOSC Portal accessible Q4 2019: Updated EOSC Catalogue of services & EOSC Portal
2018, Q2	• EOSCpilot project • EOSC-hub project • INFRAEOSC-04-2018 • INFRAEOSC-05-2018-2019 (b)	Develop **initial catalogue of datasets** accessible via the EOSC (to be enriched periodically)	Q2 2019: Initial EOSC Catalogue of datasets accessible.

Starting from	Committed resources (non-exhaustive)	Action	Milestones
2018, Q1	• EOSC -hub project • OpenAIRE-Advance project • FREYA project	Develop initial EOSC federating core including the EOSC shared resources	Q4 2019: Initial EOSC federating core in place
2019, Q1	• EOSCpilot project INFRASUPP-01-2018-2019 (b3)	Develop catalogue of interested and eligible (per Rules of Participation) data infrastructures to be federated into the EOSC and identify EOSC federate centres	Q4 2019: Registry of data infrastructures of the EOSC (initial)
2018, Q4	• INFRAEOSC-04-2018	Connection the research infrastructures identified in the ESFRI Roadmap to the EOSC. Support to this activity will be provided through cluster projects.	Q2 2020: Preliminary connection of most infrastructures and services to the EOSC

COMMISSION STAFF WORKING DOCUMENT Implementation Roadmap for the European Open Science Cloud, 3 mars 2018

Les enjeux

En résumé

Enjeux politiques au niveau européen

- Open science et science citoyenne
- Création de l'Espace européen de la recherche
- Stratégie pour un marché unique numérique en Europe
- Implication des financeurs de la recherche au niveau des Etats-membres

Évolutions des pratiques de recherche

- Changements méthodologiques
- Anticipation
- Aptitudes et compétences des chercheurs
- Formation doctorale
- Évaluation des carrières scientifiques
- Nouveaux métiers au sein des RPO
- Culture du partage dans toutes les disciplines
- Intégrité scientifique, transparence
- Responsabilité individuelle, déontologique et éthique des chercheurs
- Interdisciplinarité
- Collaborations, science citoyenne

Enjeux économiques

- Coût des infrastructures, modèle économique d'EOSC
- Valorisation des données de la recherche pour produire de la richesse
- Engagement des acteurs institutionnels

Enjeux juridiques

- Prise en compte des autres régimes juridiques comme RGPD, données sensibles, sûreté nationale
- Propriété intellectuelle liée aux données (ouverture, ré-utilisation)
- Principe de réciprocité

Enjeux politiques au niveau européen

- Open science et science citoyenne
- Création de l'Espace européen de la recherche
- Stratégie pour un marché unique numérique en Europe
- Implication des financeurs de la recherche au niveau des Etats-membres

Enjeux économiques

- Coût des infrastructures, modèle économique d'EOSC
- Valorisation des données de la recherche pour produire de la richesse
- Engagement des acteurs institutionnels

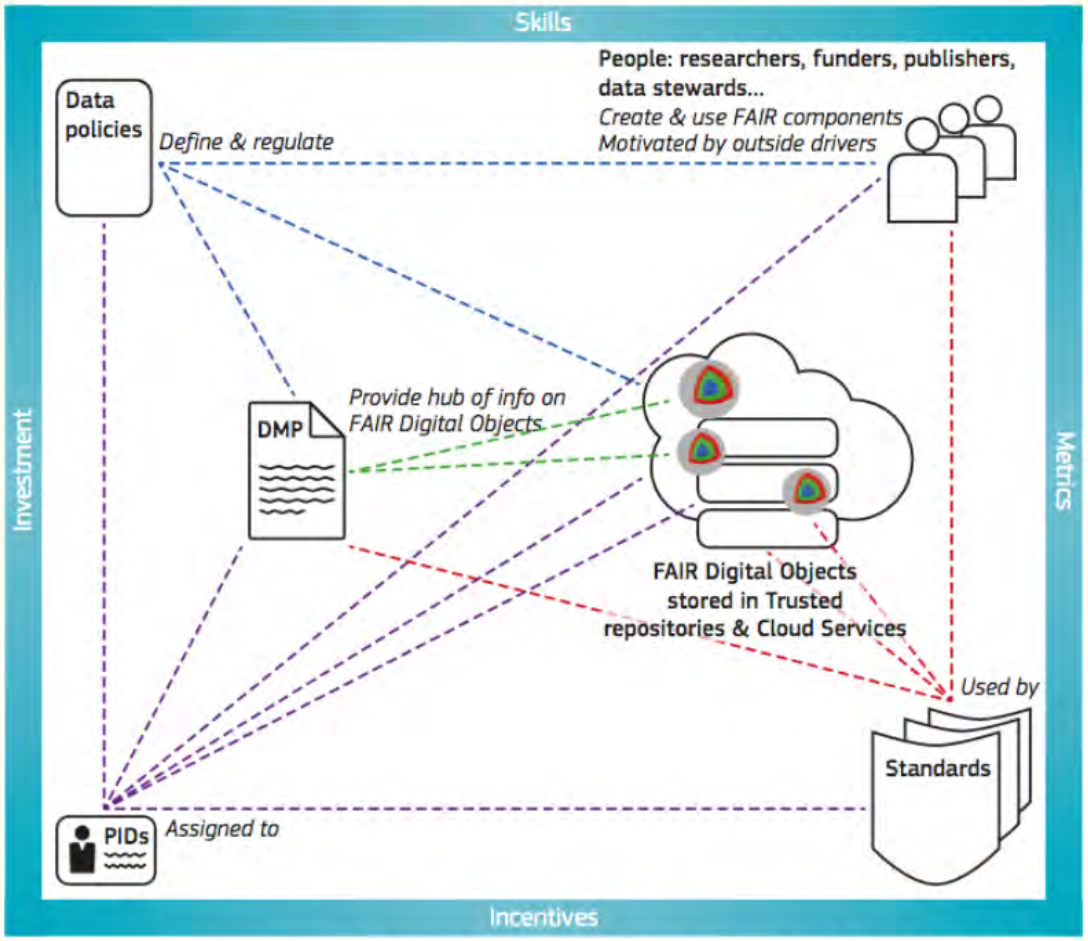
Enjeux juridiques

- Prise en compte des autres régimes juridiques comme RGPD, données sensibles, sûreté nationale
- Propriété intellectuelle liée aux données (ouverture, ré-utilisation)
- Principe de réciprocité

Évolutions des pratiques de recherche

- Changements méthodologiques
- Anticipation
- Aptitudes et compétences des chercheurs
- Formation doctorale
- Évaluation des carrières scientifiques
- Nouveaux métiers au sein des RPO
- Culture du partage dans toutes les disciplines
- Intégrité scientifique, transparence
- Responsabilité individuelle, déontologique et éthique des chercheurs
- Interdisciplinarité
- Collaborations, science citoyenne

En résumé



Source : Turning FAIR into reality - Final Report and Action Plan from the EC Expert Group on FAIR Data, 2018

Horizon Europe ?

- Accentuer l'ouverture (groupe Lamy de haut niveau)
- Planification stratégique :

Open Science practices will be mainstreamed as the new *modus operandi* for EU research and innovation. Particular focus will be placed on open access to scientific publications and research data, management of research data along the FAIR principles, development and consolidation of the European Open Science Cloud (EOSC) to provide a trusted and open virtual environment for data-driven research for all researchers and innovators, and responsibility and openness of science towards society, and vice versa. Open science will increase the quality of science and productivity of research and will accelerate the pace and uptake of innovation to deliver on societal, economic and technological challenges. It will also give Europe a global lead in research data management. Promoting and enabling open responsible research will help improve trust between science and society and the uptake of scientific evidence-based public policies and innovative solutions.

- Place importante des données de santé
- Infrastructures : soutien à EOSC, HPC, EDI

Merci !

corinne.joffre@univ-toulouse.fr