EU legal regulations and their impact on policies for EOSC

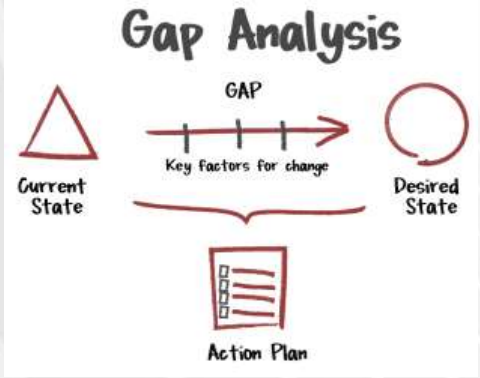
* Caterina Sganga (Scuola Superiore Sant'Anna)
* Nadina Foggetti (INFN)

Legal and policy analysis: what we did

* Comparative legal mapping (EU and national laws)
* Gap analysis
* Policy recommendations
* Legal Compliance Guidelines for Researchers: a Checklist
* What’s next in EU law, and with which impact?
* Use cases

Comparative legal mapping

* See “Legal and Policy Framework and Federation Blueprint” (https://doi.org/10.5281/zenodo.5647948)
* Collection and assessment of laws from EU, Austria, Belgium, France, Germany, Italy impacting on OA and FAIR ecosystem, i.e.
    * Copyright
    * Personal data
    * Non-personal data



IN GENERAL

* System of copyright exceptions and limitations (E&L) makes it impossible for the © system to respond to changing needs of research ecosystems and evolution of technologies
* Copyright contracts are NOT standardized and often not compatible with OA and OS principles
* Process towards open data strategy not completed yet
* Legislative fragmentation at national level

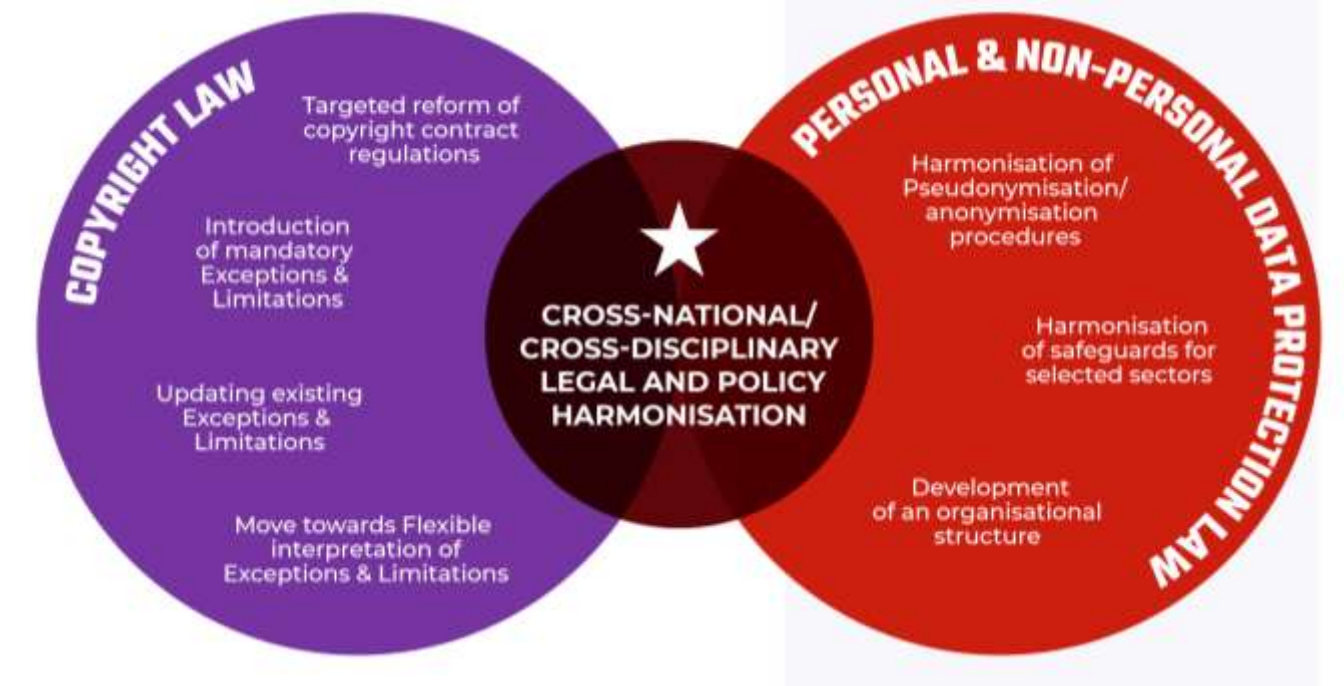
Regulatory constraints: COPYRIGHT

* Scope and enforcement of exclusive rights
* Strict reading and lack of flexibilities of L&Es
* Not full harmonization of L&Es
* Ample room left for freedom of contract, often imposing additional constraints and broadening exclusive rights
* Breadth of definition of protected works

Regulatory constraints: PERSONAL / NON-PERSONAL DATA

* National implementation of specific norms (e.g. exceptions) not enacted by similar legal sources different effectiveness among different legal systems
* National safeguards developed via different and oft-conflicting approaches
    * (e.g. tech and organizational measures, boundaries under general research purpose regime etc)
* National provisions do not rely on technical standards required to achieve interoperability missing standardisation

Policy recommendations



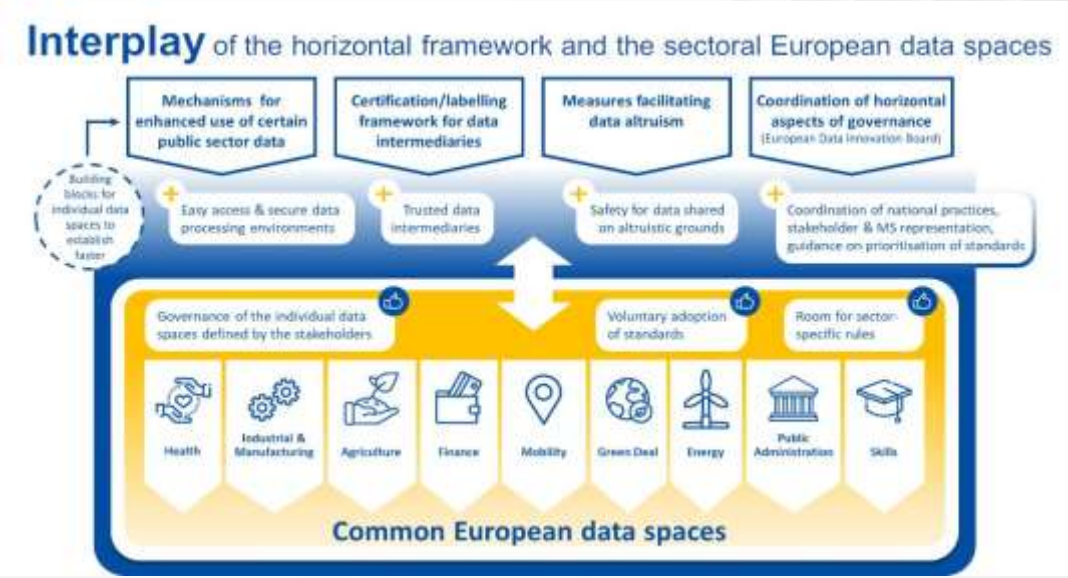
The Checklist

* See “Legal Compliance Guidelines for Researchers: a Checklist”
  * Both digital (with interactive checkboxes, https://doi.org/10.5281/zenodo.632766), and printable (https://doi.org/10.5281/zenodo.6327691)
* To guide researchers in management of research outputs vis-à-vis IP and data protection laws
* To promote best practices to achieve FAIR ecosystemt, removing unnecessary restrictions to reuse and access + facilitating convergences of national solutions

The Checklist

RESEARCH PROPOSAL 1	RESEARCH IMPLEMENTATION 2	RESEARCH REVIEW 3
Background informations, IPR, Exploitation, TTO, FAIR, DPIA.	IP Management Plan, IP FLEXIBILITY, GDPR, DMP, Findable, Accessible, Re-Usable, interoperable.	IP Management Plan, Licences and FAIR/OS/OA, Re-use of Data.

What's next?



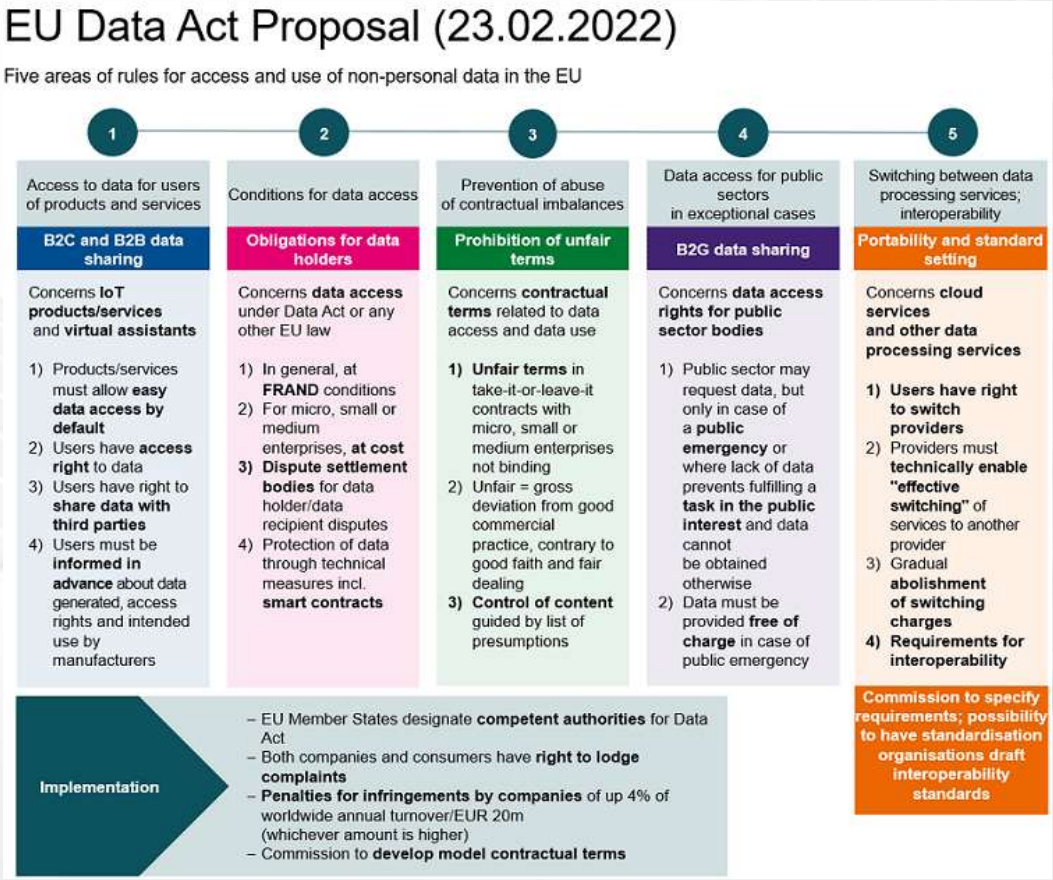
See https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020SC0295&rid=6

What’s next: the Data Governance Act

- Fill in important gaps in public sector data sharing making public sector data available for re-use, in situations where such data is subject to rights of others.
- Bridge the public/private divide + focus on private data-sharing sharing of data among businesses, against remuneration in any form
- Innovation in data governance tools
  - **Data-sharing intermediaries** allowing personal data to be used via intermediaries designed to help individuals exercise rights under GDPR
  - **Data altruism** sharing of data on altruistic grounds
  - Incentivize development of common EU data spaces

What's next?

See https://www.cms-lawnow.com/ealerts/2022/02/proposal-for-eu-data-act-adopted



EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

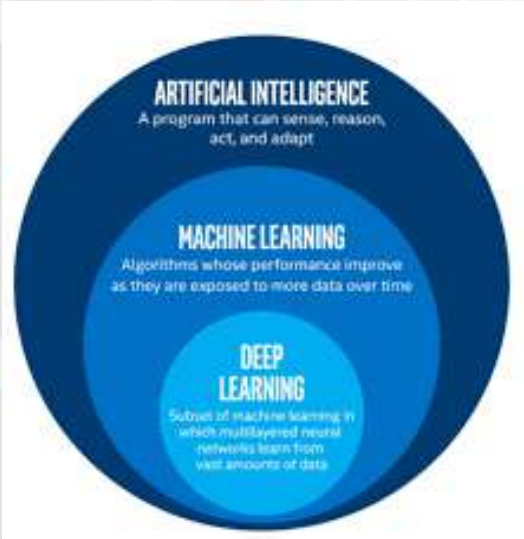
Use Case Prospective

EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.

The Role of the checklist

Checklists aim to:

* A. To guide researchers in the management of data, or more generally, of research results,
* B. Promote the adoption of best practices to achieve the findability, accessibility and interoperability of research data, focusing on removing unnecessary restrictions on reuse and open access to published products, facilitating convergence between national solutions.



Proposal

Check if information is available on the data and intellectual property rights introduced in the project.

If you are processing data belonging to particular categories, check whether or not the guarantees defined in the checklist are necessary.

If you intend to process personal and non-personal data using technologies based on artificial intelligence or machine learning techniques: legal and ethica requirement are define in Checklist

Resaerch implementation

IPR

- The definition of IP acquisition principles and policies
- Authorization processes and systems that identify IP flexibilities and make the most of them

DP

- Data Management plan (DPM)
- Innovative use of data processed policy

Research review

identify and keep updated processes and systems able to identify the flexibility of the intellectual property system

Verify compliance with license agreements by your licensees, and in particular compliance with the FAIR, Open Access and Open Science clauses.

Introduce mechanisms for verifying data retention times and provide for actions necessary to ensure anonymization and / or cancellation

identify data usage policies

Remove any access to subjects / entities / collaborators no longer authorized. Verify that you have followed all the instructions provided within the data management plan (DMP).

EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Use case 1

T.6.6 of EOSC_Pillar

EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.

Use Case 6 Eosc Pillar

* In the context of Task 6.6 of EOSC Pillar project the Use case 6, aims to explore reference data through existing computing services for the bioinformatics community (INSERM)
  The aim of this use case will be to explore the possible interactions between already available Galaxy computing services and data repositories, in order to build an integrated and interoperable service for ELIXIR and the wider Life Science user community as a whole. It will aim at fulfilling the following objectives:

- Allow frictionless **access to external data** sources from different Galaxy deployments
- Facilitate the deployment of Galaxy instances **close to the data**
- Provide **coherency** between different existing Galaxy deployments
- Ensure health **data security** requirements are met throughout the process

Partners involved

* CNR (IT)
* INFN (IT)
* INSERM (FR)

Laniakea

Laniakea

https://laniakea-elixir-it.github.io

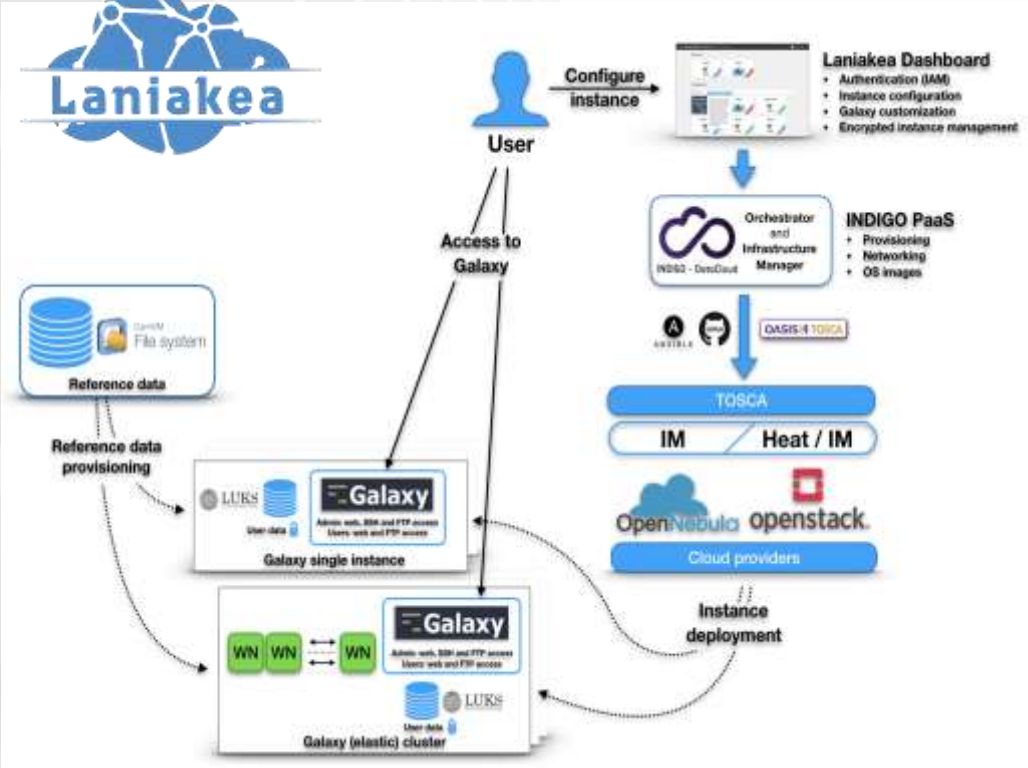
LANIAKEA is a cloud Galaxy instance provider, based on INDIGO-DataCloud software catalogue. Its architecture automates the creation of Galaxy-based virtualized environments exploiting the software catalogue provided by the INDIGO-DataCloud project.

No need for the end user to know the underlying infrastructure.

No need for maintenance of the hardware and software infrastructure.

(*) The Laniakea Supercluster (Laniakea; also called Local Supercluster or Local SCl or sometimes Lenakaeia) is the galaxy supercluster that is home to the Milky Way and approximately 100,000 other nearby galaxies [Wikipedia].

Laniakea architecture

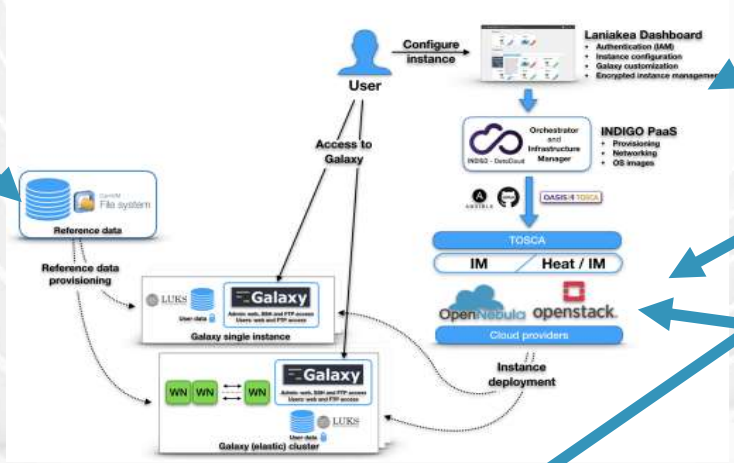


- **Dashboard -** User friendly access to configuration and and launch of a Galaxy instance.
- **IAM -** Authentication and Authorization system.
- **INDIGO PaaS -** Galaxy automatic deployment.
- **Cloud Providers -** (INFN) ReCaS-Bari and others.
- **Persistent storage** - With/without encryption.
- **Reference data availability** - With CERN-VM FileSystem.
- **CLUES -** Elasticity manager.

Legal issue and the application of Cheklist

Identify the conditions / and data protection
based on internal and international regulations

France LS vs Italian LS
The pseudonimization requirements:
IT-> simple pseudonimization
FR-> Entity for pseudonimization



Authentication system legal Compliant

Connection with DPO and TTO

Measure needed to use genetic and genomic data

Data retention conditions
Process of reusability of data (pesudonimization e/o aninimization

Plan for the use of data compliance with EU and National law IPR

Repository choice: compliance with National and International Legal Framework

Legal Framework for the use and re-use of health data for scientific purposes
DOI: 10.5281/zenodo.6334878

EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Local EGA

USE case 2

EUROPEAN GENOME-PHENOME ARCHIVE

EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.

The European Genome-phenome Archive

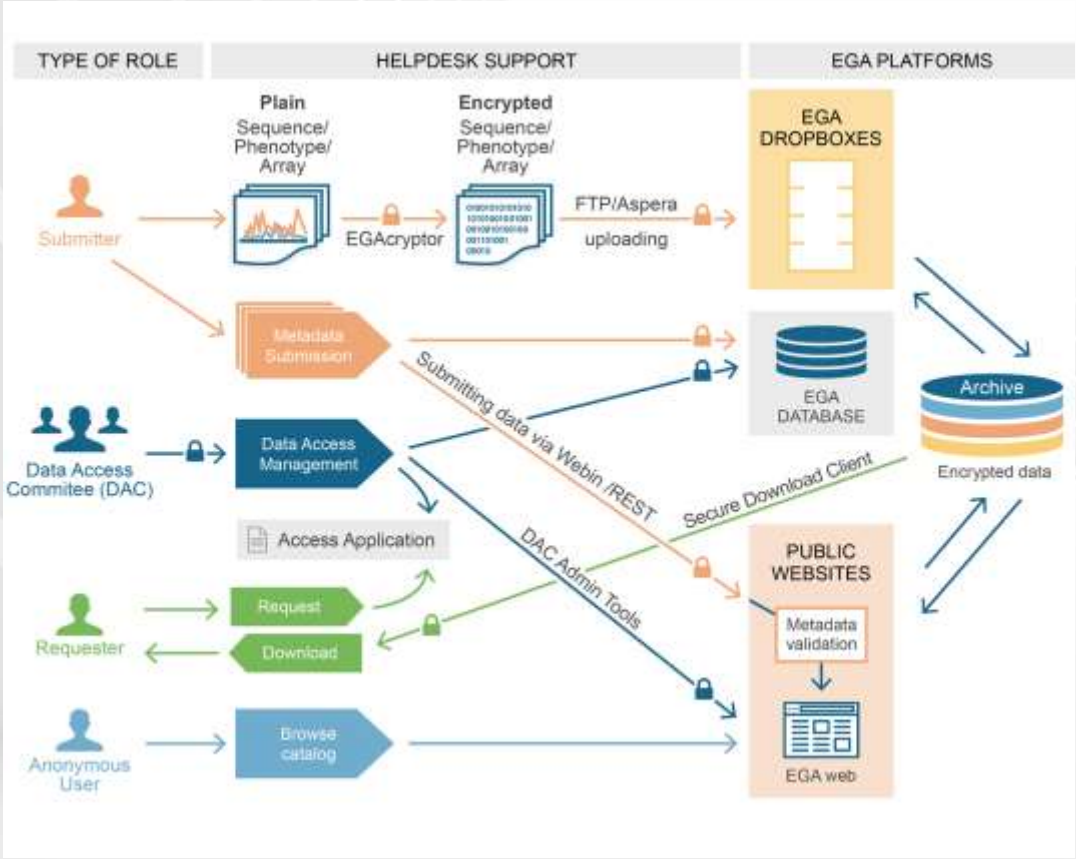
EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.

The European Genome-phenome Archive (EGA) is a service for permanent archiving and sharing of all types of personally identifiable genetic and phenotypic data resulting from biomedical research projects.

Data at EGA was collected from individuals whose consent agreements authorise data release only for specific research use to bona fide researchers.

Strict protocols govern how information is managed, stored and distributed by the EGA project.

The European Genome-phenome Archive



Studies and datasets can be browsed by anonymous users.

Data access committee is responsible for approving access to single of multiple datasets.

Data are encrypted. Trusted users exploit a user-specific key to decrypt data.

Local EGA

It aims at solving the issue where sensitive data cannot move across borders (cf to GDPR), while public metadata can. Files will be stored encrypted in the Local EGAs located in different countries, while public metadata stays at Central EGA.

1. Submitters upload encrypted files into a Local EGA inbox, located in the relevant country.
2. Encrypted files are moved from to long-term storage, and information are saved in Local EGA database.
3. In the process, each ingested file obtain an Accession ID, which identifies it uniquely across the EGA.
4. The distribution system allows requesters to access securely the encrypted files in the long-term storage, using the accession id, if permissions are granted by a Data Access Commitee (DAC).

Federete Lega

- The EGA is currently transitioning from a centralised resource managed by EMBL-EBI (Hinxton, UK) and CRG (Barcelona, ELIXIR Spain, with key support of the Barcelona Supercomputing Centre) to a federated node model. The Federated EGA is designed to support national data management requirements for genomic and clinical data collected from their citizens as part of healthcare or biomedical research projects.

Data in Life Science

Genomic data are distributed across several sequencing centres and/or IT infrastructures for LEGA Use Case

Trasnational legal context:

- Define the legal framework IPR and Data Protection
- Define the requirements for data acquisition
- Define the requirements for data processing
- Define the requirements for storage of data

Key issue

Genomic data are sequenced in national sites

The result may be accessible from different countries and users

Definition of the requirement for use of results

Application of Checklist to Legal use Case

* Definition of the the legal regime applicable to each research product, with particular regard to territoriality and cross-border activities .

* **Liaise with Technology Transfer Office (TTO) or offices in charge of legal matters, to lay down internal processes for the protection and management of Intellectual Capital (IC) stemming from the project**

* Make sure that your TTO or other office is aware of FAIR principles, and OA/OS best practices.

* Data acquisition requirements: Consens of patients; use of public dataset, use of research output dataset etc.

* Use of personal data extra EU: study of the conditions applicable for the use

* Technical and organisational measures required to protect data flows in terms of Availability, confidentiality, integrity.

Future prospective

- Implementation of the Guidelinees and Cheklist in the Italian landscape:
  - ICDI
    - Gap Analysis for the complicance with Italian Law (IPR and Data Protection Law)
      - Definition and Italian Cheklist and Policy in order to promote protection and openess!