

Ett ramverk för interoperabilitet inom European Open Science Cloud (EOSC)

(16 Mars 2021)

Magnus Eriksson, Enheten för Data som Strategisk Resurs, Vetenskapsrådet, medlem av expertgruppen EOSC FAIR WG och arbetsgruppen Interoperability taskforce.

Syfte

Detta dokument syftar till att ge en sammanfattande beskrivning av arbetet i EOSC:s arbetsgrupp för Interoperabilitet (EOSC FAIR WG - Interoperability Taskforce). Arbetsgruppens resultat kan få inverkan på hur arbete kring interoperabilitet organiseras i olika projekt som VR är involverade i och kan komma att användas som vägledning i Vetenskapsrådets engagemang i EOSC specifikt men också allmänt runt VRs arbete runt öppen tillgång till data.

Inledning

Vikten av återanvändning och delning av data och tjänster för att forskare ska kunna samverka effektivt och nå resultat som är viktiga för samhället har blivit väldigt tydligt under den pågående pandemin.

Behovet av tillgång till data av hög kvalitet som är lätt att hitta, få åtkomst till (om behörighetskrav uppfylls), utbytas och användas tillsammans med andra data samt återanvändas för att lösa samhällsutmaningarna har ökat. Under det senaste året har detta synliggjorts i arbetet mot Covid-19.

Förmågan att på ett effektivt sätt utbyta data och kunna använda data från olika källor tillsammans kallas *Interoperabilitet*. Om vi vill att forskare ska kunna lägga mindre tid på att söka, utreda och bearbeta datakällor så att de går att använda och istället kunna öka fokus på sin forskning krävs god interoperabilitet.

För att skapa en "virtuell miljö" där forskare kan använda samt dela data och tjänster krävs en infrastruktur som stödjer detta. Den infrastrukturen behöver byggas gemensamt, baserat på gemensam policy och riktlinjer för hur data ska hanteras så att inte alla gör på olika sätt (vilket skulle göra det svårare att utbyta data). Om det vi utvecklar dessutom använder gemensamma centrala komponenter på ett överenskommet sätt förbättras interoperabiliteten ytterligare.

En sådan "virtuell miljö" skapas i European Open Science Cloud genom att skapa en federation av tjänster, dvs ett nätverk av system som använder gemensamma komponenter, kommunicerar och används på ett överenskommet sätt.

Då behoven ibland kan skilja sig åt mellan vetenskapliga områden kan även riktlinjer och komponenter behöva anpassas gemensamt inom respektive vetenskapligt område. Både de komponenter som kan användas generellt och de som behöver anpassas behöver följa samma "mall" så att inte interoperabiliteten påverkas negativt. Denna "mall" kallas ramverk och den detaljerade "ritningen" som används vid utveckling kallas "referensarkitektur".

Arbetet med interoperabilitet samt överenskommelser om gemensamma riktlinjer och användning av gemensamma centrala komponenter bör följa hela datats livscykel för att öka graden av interoperabilitet på ett effektivt sätt. Med anledning av detta är det av största vikt att datahantering baserat på gemensamma riktlinjer implementeras nationellt och internationellt, även hos organisationer med viktiga datakällor för forskning, exempelvis i sjukvårdens regioner och hos registerhållande myndigheter.

I rapporten EOSC Interoperability Framework beskriver expertgruppen "EOSC FAIR WG – Interoperability taskforce" de behov av interoperabilitet som ett antal forskningsinfrastrukturer uttryckt samt hur EOSC kan gå vidare med att realisera detta i ett ramverk för interoperabilitet och med stöd av en referensarkitektur.

EOSC -en federation av tjänster

European Open Science Cloud (EOSC) är ett initiativ som syftar till att etablera en federation av tjänster för att tillgängliggöra forskningsdata och forskningsstödande tjänster. En sådan federation kan liknas vid ett nätverk där ett antal datorsystem samverkar för att kunna använda varandras resurser. Dessa resurser kan till exempel innebära funktioner för tillgängliggörande av forskningsdata för andra system eller andra funktioner som behövs för att nätverket ska fungera.

Datorsystemen och deras funktioner benämns i detta fall tjänster och kan vara av olika slag. En tjänst kan till exempel innebära möjligheten att läsa forskningsdata för de som har behörighet att göra så. En annan tjänst kan möjliggöra länkning till begrepp som definierar data på ett gemensamt sätt medan en ytterligare tjänst kan innebära tillgång till beräkningsresurser på en superdator.

En federation av tjänster har således ett stort behov av att kunna utbyta data och använda olika typer av tjänster på ett gemensamt och effektivt sätt. Detta är något som är mycket komplext och denna komplexitet behöver hanteras på ett strukturerat sätt för att utbyte och användning ska möjliggöras på ett kostnadseffektivt sätt.

Ramverk skapar interoperabilitet

Det första steget för att hantera komplexiteten i en federation har varit att dela in de olika frågorna som tjänsteleverantörer och användare av tjänster behöver komma överens om för att vara interoperabla, dvs att fungera tillsammans och kunna utbyta data och tjänster på ett effektivt sätt och inom olika områden. Detta görs genom att beskriva indelningen i så kallade ramverk.

De europeiska myndigheternas ramverk

Det så kallade European Interoperability Framework (EIF)¹ har möjliggjort utbyte av data och tjänster för myndigheter inom EU. EIF har identifierat fyra olika områden inom vilka utmaningar behöver lösas och gemensamma riktlinjer behöver tas fram. Dessa kan delas in i 1) legala; -2) organisatoriska;- 3) semantiska respektive tekniska frågor/områden.

Myndigheter inom EU har dock en mer likartad styrning, verksamhet och mognadsgrad än det stora antal forskningsprojekt som bedrivs inom olika vetenskapliga domäner. Detta innebär att mängden, och typen av, frågor där leverantörer av tjänster och dess användare behöver komma överens för att vara interoperabla växer avsevärt och att komplexiteten ökar inom forskningsdataområdet.

Arbetsgruppen "EOSC FAIR WG Interoperability taskforce" utsågs av den tidigare EOSC Executive Board 2019, och fick i uppdrag att sammanställa ett ramverk motsvarande myndigheternas EIF men för EOSC och forskningsdataområdet. Arbetsgruppen har varit verksam mellan 2019-2020 och bestått av experter inom Ontologier (avdelning inom AI), Superdatormiljöer, Life Science, Forskningsdatainfrastrukturer inom hälsodata, maskininlärning, forskningsdatastöd och en Forskningsfinansiär. Arbetsgruppen presenterade sitt förslag på ramverk för interoperabilitet inom forskningsdataområdet i rapporten EOSC Interoperability Framework i februari, 2021. Nedan följer en sammanfattning av processen med att ta fram förslaget och medföljande rekommendationer.

¹ https://ec.europa.eu/isa2/eif_en

Ett ramverk för interoperabilitet mellan datasystem och tjänster inom EOSC

För att få en bild av vilka frågor som är mest centrala för att uppnå interoperabilitet och vilka hinder som finns gjorde arbetsgruppen intervjuer med europeiska forskningsinfrastrukturer som arbetar med att tillhandahålla stöd och tjänster till forskare idag samt med olika projekt.

Med detta som utgångspunkt identifierades de generella principer som skulle styra framtagandet av EOSC:s ramverk för att uppnå interoperabilitet inom forskningsområdet. Det ramverk som tagits fram för myndigheter inom EU (EIF) blev startpunkten för arbetet inom forskningsområdet då de områden inom vilka olika frågeställningar kunde samlas var snarlika, dvs inom de legala, organisatoriska, semantiska och tekniska områdena.

För varje område dokumenterades de problem, behov och utmaningar som kommit fram under intervjuerna samt från relevanta litteraturstudier. Ett antal rekommendationer/riktlinjer på övergripande nivå formulerades för att användas vid utveckling av EOSC:s ramverk för interoperabilitet.

En första version av ett ramverk för interoperabilitet inom EOSC (EOSC Interoperability Framework; EOSC IF) publicerades som en delrapport och lades ut för konsultation (2020-05-17) så att forskare, infrastrukturer och intressegrupper inom forskningsdataområdet kunde lämna kommentarer. Ramverket presenterades även på workshops för att få återkoppling.

Ett stort antal kommentarer togs emot i samband med konsultationerna och gemensamt för flertalet av dessa var ett önskemål om en mer detaljerad och tekniskt relevant beskrivning av de olika områdena och vilka delar som de bestod av mer i detalj. Kommentarererna togs omhand i det vidare arbetet genom att en mer teknisk beskrivning av ramverket togs fram i form av en referensarkitektur.

En referensarkitektur för interoperabilitet inom EOSC

Referensarkitekturen utgör en mer teknisk beskrivning av EOSC Interoperability Framework och beskriver en struktur för att vägleda arbetet med att nå god interoperabilitet mellan datasystem och tjänster inom EOSC.

Referensarkitekturen beskriver de olika delar där gemensamma överenskommelser om riktlinjer, standards, verktyg, metoder, kod m.m. behöver utarbetas i detalj. De olika delarna kallas för byggnadsblock och dessa grupperas i tekniska beskrivningar av ramverket beroende på vilken funktionalitet de är en del av att implementera.

Utgångspunkten var här de behov som framkommit i det tidigare arbetet med EOSC Interoperability Framework samt de som EOSC Sustainability Work Group formulerat och benämnt Minimal Viable EOSC (MVE) i rapporten "[FAIR Lady](#)"². I FAIR Lady rapporten utforskas de krav som ställs på EOSC för att det ska vara långsiktigt hållbart från många olika perspektiv.

Kommentarer från konsultationerna, gällande EOSC Interoperability Framework, pekade på behovet av att referensarkitekturen för forskningsdataområdet behövde beakta de Europeiska myndigheternas referensarkitektur för interoperabilitet (EIRA). Detta för att ta tillvara erfarenheterna från det europeiska arbetet inom området. Bättre interoperabilitet vid utbyte av data mellan forskningsprojekt och olika typer av myndigheter är naturligtvis också en positiv effekt av detta. Då arbetsgruppen gjort motsvarande på ramverksnivå fanns förutsättningar för att göra detsamma med den mer tekniska referensarkitekturen.

²DOI: 10.2777/870770

Som tidigare nämnts är forskningsdataområdet samt system och tjänster inom detta betydligt mer omfattande och diversifierat än de europeiska myndigheternas. Arbetsgruppen tog därför fram en utvidgning (Extension) av de europeiska myndigheternas referensarkitektur (EIRA) enligt de riktlinjer som EIRA föreskrivit. Utvidgningen utgör *EOSC Interoperability Reference Architecture*.

Referensarkitekturen utvecklades sedan vidare för att möjliggöra återanvändning av dess delar vid utveckling av såväl stödande infrastruktur som vid utveckling av de tjänster som levererar innehållet inom EOSC-federationen. Möjligheten till flexibilitet och återanvändning av de gemensamma riktlinjer, koder, verktyg mm som samordnas inom de olika "byggnadsblocken" är extra viktigt inom forskningsdataområdet då det är mycket mer diversifierat än myndigheternas ekosystem.

Översikt bilden nedan (Bild 1) sammanfattar hur de fyra perspektiven på interoperabilitet från EIF lägger grunden för EOSC Interoperability Reference Architecture.

Den tekniska vyn delas in i tre delar i referensarkitekturen för att förbättra förutsättningarna för flexibilitet och återanvändning av ingående delar. Byggnadsblock som utgör ett ramverk för licenshantering kan exempelvis användas i infrastruktur för dataåtkomst men även återanvändas i infrastruktur för utbildning eller metadatarepositorium.

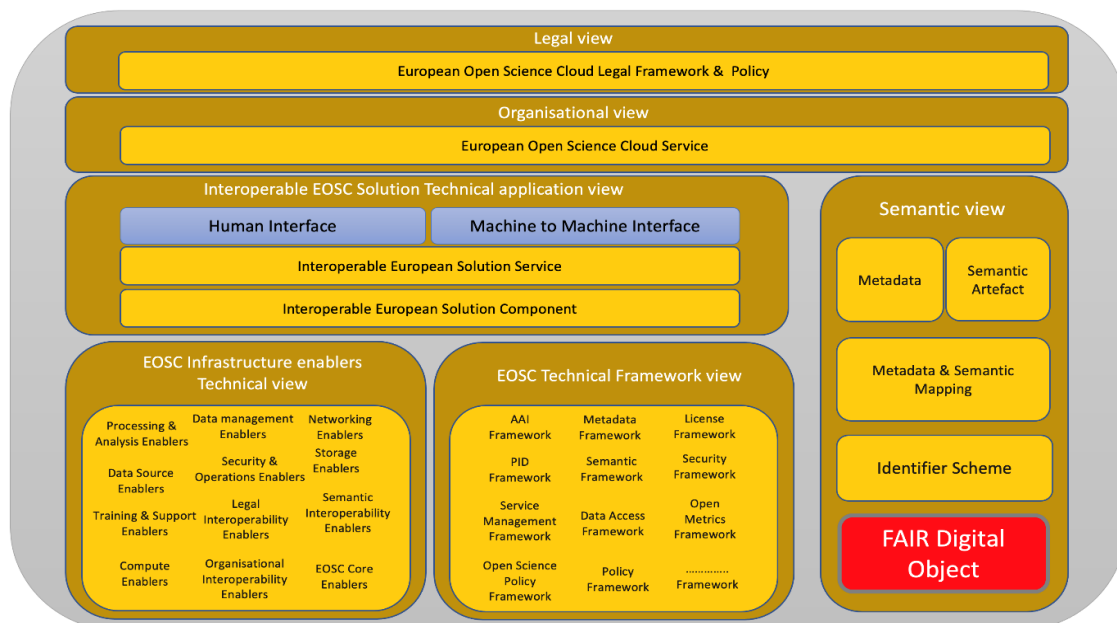


Bild 1. EOSC Interoperability Reference Architecture

Både EOSC infrastruktur och övriga tjänster tillgängliggör gränssnitt för människor och/eller maskiner och data som utbyts utgörs av digitala objekt som är FAIR³ (Findable, Accessible, Interoperable, Reusable) och beskrivs av metadata och semantik. I exemplet ovan är således licensen, datasetet, variabeln, utbildningsfilmen mm alla digitala objekt som beskrivs med metadata och semantik och tilldelas en unik identifierare.

Vad de olika byggnadsblocken ska innehålla, dvs vilka riktlinjer, standarder, verktyg m.m. som respektive vetenskapliga område ska förhålla sig till behöver fastställas och dokumenteras av områdena själva och de forskningsdatanätverk som de samverkar i och stöds av. Eftersom olika

³ Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, Article number: 160018. DOI 10.1038/sdata.2016.18. <https://www.nature.com/articles/sdata201618>

vetenskapliga områden har olika behov men också olika stor erfarenhet av FAIR datahantering kommer därför byggnadsblockens innehåll att kunna skilja sig åt.

Konsultationernas avtryck i referensarkitekturen

När arbetsgruppen tog fram EOSC:s referensarkitektur för interoperabilitet var strävan att möta de behov som uttryckts i konsultationerna genom att designa

- en referensarkitektur som stödjer utveckling av graden av FAIR inom olika vetenskapliga områden med hänsyn till att deras förutsättningar skiljer sig åt;
- en referensarkitektur som är designad för att kunna utvidgas och utvecklas baserat på återkoppling från olika communitys och vetenskapliga domäner;
- en referensarkitektur som är en utvidgning (extension) av myndigheternas referensarkitektur för att skapa så goda förutsättningar som möjligt för interoperabilitet mellan myndigheternas datasystem och tjänster och EOSC.

Rekommendationer

Arbetsgruppens rekommendationer finns publicerade i [rapporten "EOSC interoperability framework - Report from the EOSC Executive Board Working Groups FAIR and Architecture"](#)⁴ och summeras på engelska i bilaga A nedan.

Rekommendationerna är grupperade efter de fyra olika områden som återfinns i European Interoperability Framework⁵ (legala, organisatoriska, semantiska och tekniska området) och riktar sig mot olika målgrupp beroende på område.

De roller som arbetar med

- legala frågor är målgruppen för rekommendationer inom det legala området
- gemensamma processer och praxis är målgruppen för rekommendationer inom det organisatoriska området
- informatik, semantisk annotering och mappning är målgruppen för rekommendationerna inom det semantiska området
- utveckling och användning av teknisk infrastruktur och tjänster är målgruppen inom det tekniska området

Rekommendationerna på övergripande nivå i EOSC Interoperability Framework konkretiseras och detaljeras vidare för varje område i EOSC Interoperability Reference Architecture på s35.

Referensarkitekturen riktar sig till kommande leverantörer av tjänster inom EOSC, som vill tillgängliggöra olika typer av resurser, samt de roller och aktörer som vill konsumera dessa. Ramverket är centralt oavsett om tjänsterna avser stödjande EOSC infrastruktur som behövs för att få federationen att fungera eller tjänster som levererar resurser inom EOSC.

Vidare arbete

Utöver de rekommendationer som arbetsgruppen publicerat, se ovan, så bör arbetet fortsätta genom att

⁴ "EOSC interoperability framework - Report from the EOSC Executive Board Working Groups FAIR and Architecture" - <https://op.europa.eu/en/publication-detail/-/publication/d787ea54-6a87-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-190308283>

⁵ "The European Interoperability Framework (EIF)" - https://ec.europa.eu/isa2/eif_en

- **Koordinera framtagandet av detaljerade beskrivningar för byggnadsblocken inom respektive vetenskapligt område** (och för områdesövergripande sådana). Detta behöver utföras av representanter från de vetenskapliga områdena själva med stöd från de forskningsstödande resurser som stöttar dem och de communitys de verkar inom. I många fall finns redan de standarder, metoder, verktyg m.m. som behövs men de behöver beskrivas och förankras.
- **Etablera en styrningsstruktur och förvaltning** av ramverket för att vägleda, samordna och organisera vidare arbete.

Bilaga A – Rekommendationer från EOSC FAIR WG Interoperability taskforce⁶

Layer	Recommendation
Technical	<ul style="list-style-type: none"> ● Open Specifications for EOSC Services. ● A common security and privacy framework (including Authorisation and Authentication Infrastructure). ● Easy-to-understand Service-Level Agreements for all EOSC resource providers. ● Easy access to data sources available in different formats. ● Coarse-grained and fine-grained dataset (and other research object) search tools. ● A clear EOSC PID policy.
Semantic	<ul style="list-style-type: none"> ● Clear and precise, publicly-available definitions for all concepts, metadata and data schemas. ● Semantic artefacts preferably with open licenses. ● Associated documentation for semantic artefacts. ● Repositories of semantic artefacts, rules with a clear governance framework. ● A minimum metadata model (and crosswalks) to ease discovery over existing federated research data and metadata. ● Extensibility options to allow for disciplinary metadata. ● Clear protocols and building blocks for the federation/harvesting of semantic artefacts catalogues.
Organisational	<ul style="list-style-type: none"> ● Interoperability-focused rules of participation recommendations. ● Usage recommendations of standardised data formats and/or vocabularies, and with their corresponding metadata. ● A clear management of permanent organisation names and functions.
Legal	<ul style="list-style-type: none"> ● Standardised human and machine-readable licenses, with a centralised source of knowledge and support on copyright and licenses. ● Permissive licenses for metadata (and preferably for data, whenever possible). And CC0 preferred over CC BY 4.0. ● Identification of different parts of a dataset with different licenses. ● Clearly marked instances of expired or inexistent copyright, as well as for orphan data. ● A clear list of EOSC-recommended licenses and their compatibility with Member States' recommended licenses. ● Tracking of license evolution over time for datasets. ● Harmonised policy and guidance to dealing with cases where patent filing or trade secrets may be compromised by disclosure. ● GDPR-compliance for personal data. ● Additional restrictions on access and use of data only applied in cases of applicable legislation or legitimate reasons. ● Harmonised terms of use across repositories ● Alignment between Member States national legislations and EOSC.

⁶ doi: 10.2777/620649