Digital Infrastructure for Research 2016 in Krakow 報告

国立情報学研究所　山地一禎、船守美穂

研究データ利活用協議会　2016年10月3日

DI4Rとは

- The DI4R 2016 conference is co-organised by EGI, EUDAT, GÉANT, OpenAIRE and RDA Europe.
  - EGI
    - Publicly-funded federation of over 300 data and computing data centres spread across Europe and worldwide.
    - EGI provides access to over 650,000 logical CPUs and 500 PB of disk and tape storage.
  - PRACE
    - Pan-European supercomputing infrastructure
    - 4 PRACE members (BSC: Spain, CINECA: Italy, GCS: Germany, GENCI: France) with total funding of €400M for the initial PRACE systems.
  - EUDAT
    - Collaborative Pan-European infrastructure providing research data services, training and consultancy for Researcher, Research Communities and Research Infrastructure & Data Center.
  - GÉANT
    - Pan-European data network for the research and education community
    - GÉANT connects 50 million users in over 10,000 institutions
  - OpenAIRE
    - Discovery service for open access and open research data
    - Search in 17,214,310 publications 28,243 datasets from 5,578 repositories and OA journals
  - RDA Europe

European Open Science Cloudの背景

- With the adoption of the Digital Single Markets strategy on 6 May 2015, the Commission announced the launch of a cloud for research data – the 'research open science cloud'. The 'European Open Science Cloud' aims to create a trusted environment for hosting and processing research data to support EU science in its global leading role.
- REPORT on 'Towards a Digital Single Market Act',
  4.3.2. e-government: Is concerned that cloud infrastructures for researchers and universities are fragmented; calls on the Commission, in cooperation with all relevant stakeholders, to set up an action plan to lead to the establishment of the European Open Science Cloud by the end of 2016, which **should seamlessly integrate existing networks, data and high-performance computing systems and e-infrastructure services across scientific fields**, within a framework of shared policies, standards and investments…

（出典）欧州委員会 研究・イノベーション総局 "European Open Science Cloud"
http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

プログラム例

- http://www.digitalinfrastructures.eu/programme
- Opening Plenary
  - Augusto Burgueño Arjona, "The European Open Science Cloud in 2017 and beyond"
  - Karlheinz Meier, "The brain, the universe and the need for integrated infrastructures“
- Day1 Morning
  - Community Challenges I
  - Towards interoperable einfrastructures
  - Building interoperable AAI for researchers
  - Data discovery
  - EGI Federated Cloud for developers
- Day 1 Afternoon 1
  - Community Challenges II
  - Solutions for federated service management
  - User experience of AAI
  - Data repositories for research
  - Making Use of the Linked Open Data Services for OpenAIRE: Querying Data about Research Results, Persons, Projects and Organisations
- Day 1 Afternoon 2
  - Serving the 99% of users
  - Community services
  - Cloud procurement
  - FAIR persistent identifiers
  - Visualising Operational Informatics Data using R

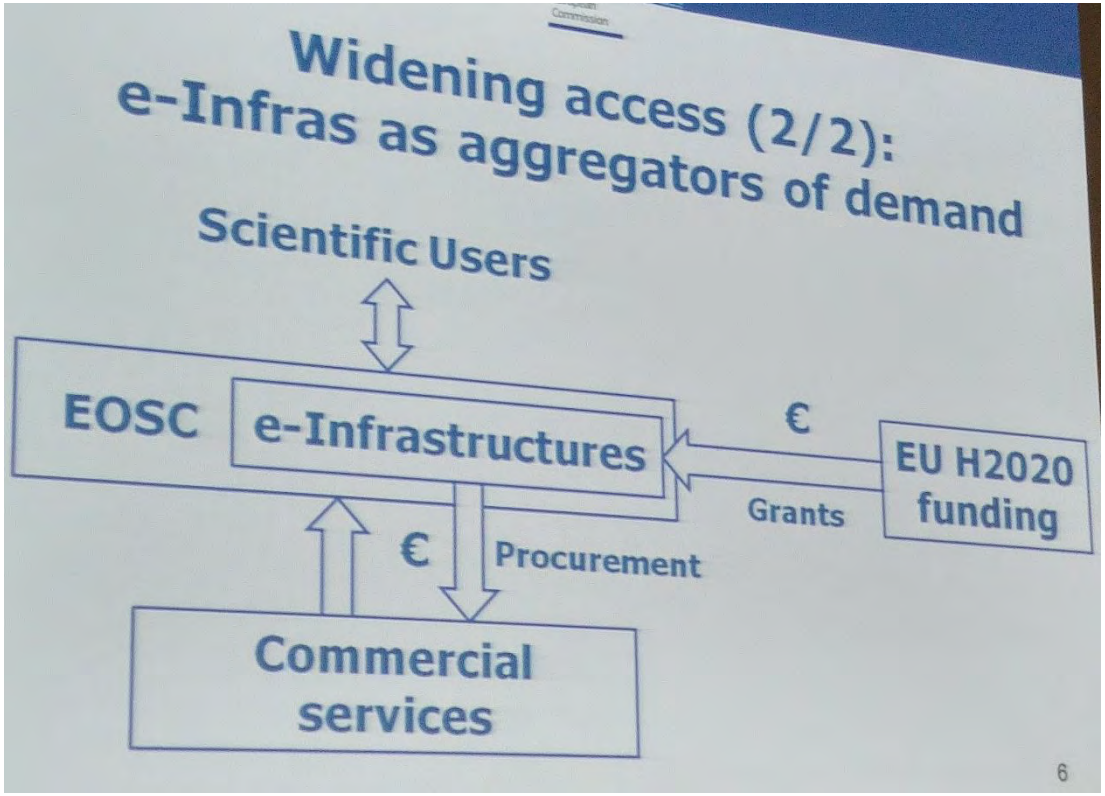
Opening Plenary：EOSC関連公募テーマ

Augusto Burgueño Arjona, "The European Open Science Cloud in 2017 and beyond“

c.f. EUDAT: FP7 16M, H2020 19M　OpenAIRE: H2020 13M

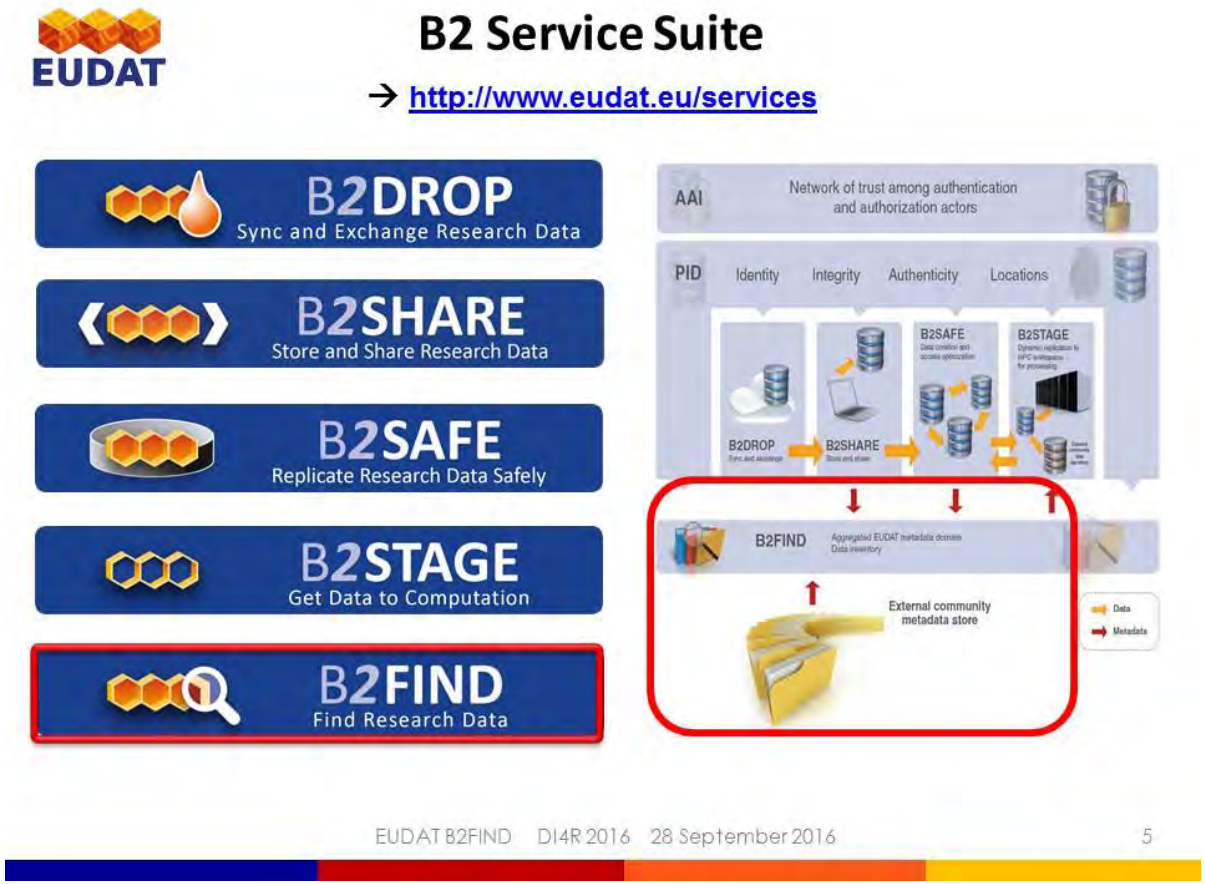
Opening Plenary：商用調達と独自開発

Augusto Burgueño Arjona, "The European Open Science Cloud in 2017 and beyond“



- 最も効率的に準備ができるように、商用を積極的に調達していく。
- H2020のグラントは内部での開発だけに使うのではなく、商用の調達にも使う。
- 商用の製品が仕様として十分でない場合はscientific useにre-package（or add-value）して提供する。
- 研究者ごとに買うのか、PIの機関で調達するかについては今後詰めていく。

EUDAT: B2FINDの位置づけ



（出典）DI4R資料 Widmann, H. “EUDAT B2FIND :A Cross-Discipline Metadata Service and Discovery Portal”
http://www.digitalinfrastructures.eu/content/eudat-b2find-cross-discipline-metadata-service-and-discovery-portal

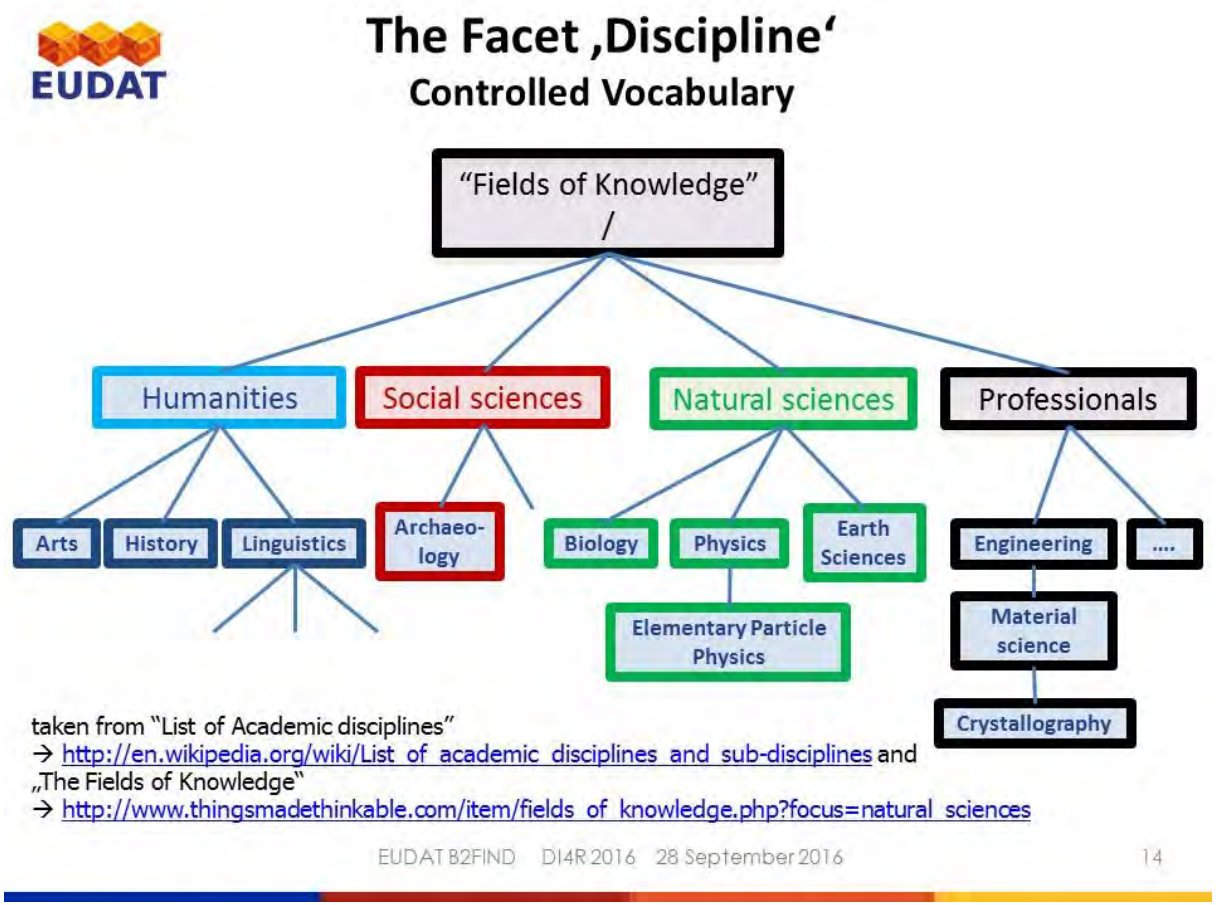
EUDAT: Metadata Schema

B2FIND MD Schema (extract)

Metadata Type	B2FIND Field name	Allowed values	Semantic definition	Level of Obligation	Occurenc	
General information	Title	Free text (unicode)	A name or title a resource is known	Mandatory	1	
	Description	Free text	Additional info	Recommended	0-1	
Data Access	Source	Valid URL or URN	Unique link to data resource	Mandatory (1)	0-1	1-3
	PID	Persistent Identifier	+ persistent and resolvable		0-1	
	DOI	Digital Object Identifier	+ citable		0-1	
Provenance data	Creator	';'-sep. list of names	Main researchers involved in data prod.	Recommended	0-n	
	Discipline	List of values from CV	Field of research (Controlled Vocab)	Recommended	0-n	
	Publication Year	YYYY	The year data are published	Recommended	1	
Formal data	Temporal Coverage	Interval of 2 DTimes [Begin, End]	The temporal limits of a date-time	Optional	1-n	
	Spatial Coverage	Spatial box or point [[minlat,minlon...]]	The spatial limits of a place.	Optional	1-n	

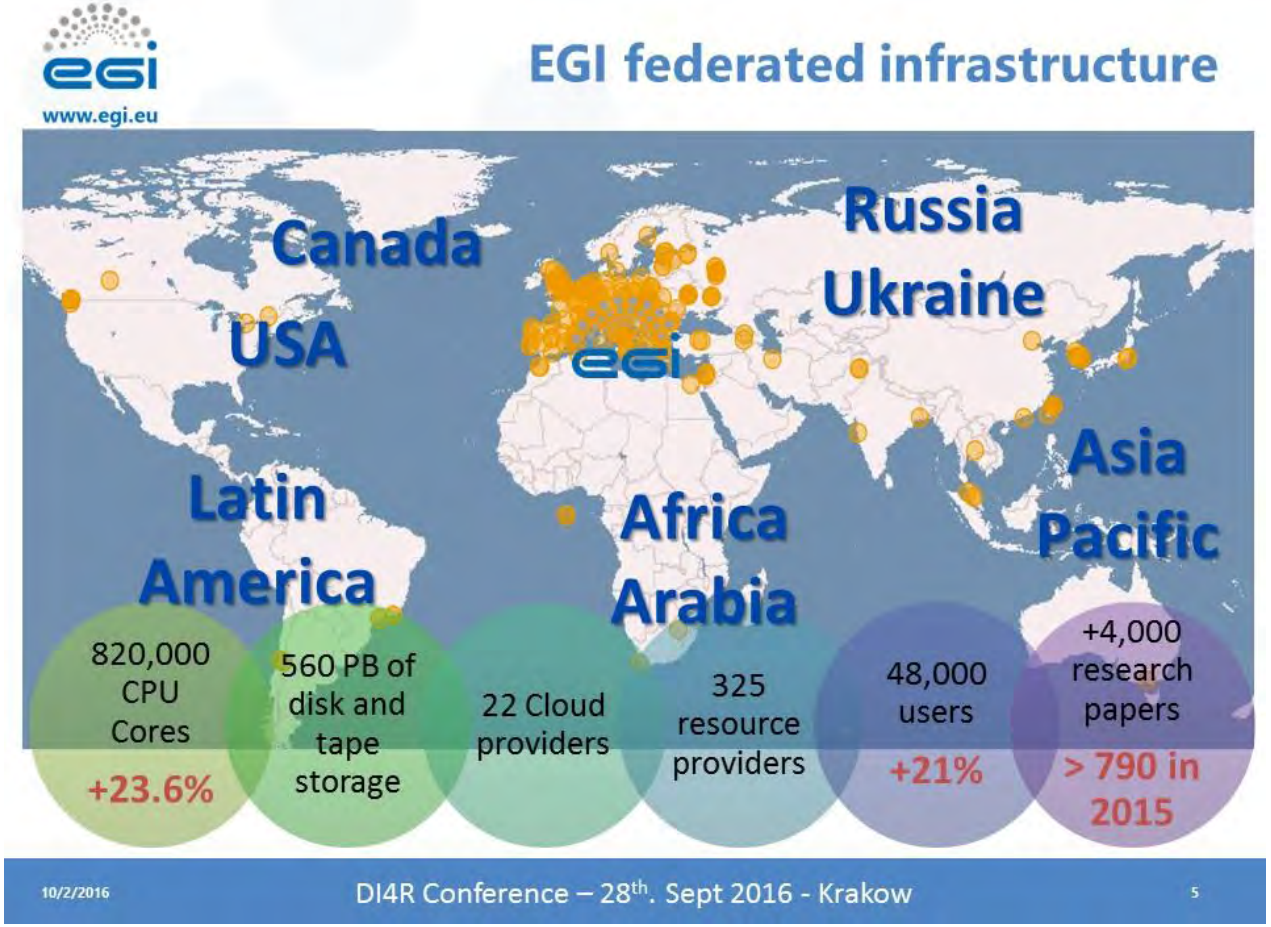
（出典）DI4R資料 Widmann, H. "EUDAT B2FIND :A Cross-Discipline Metadata Service and Discovery Portal"
http://www.digitalinfrastructures.eu/content/eudat-b2find-cross-discipline-metadata-service-and-discovery-portal

EUDAT: Discipline Controlled Vocabulary



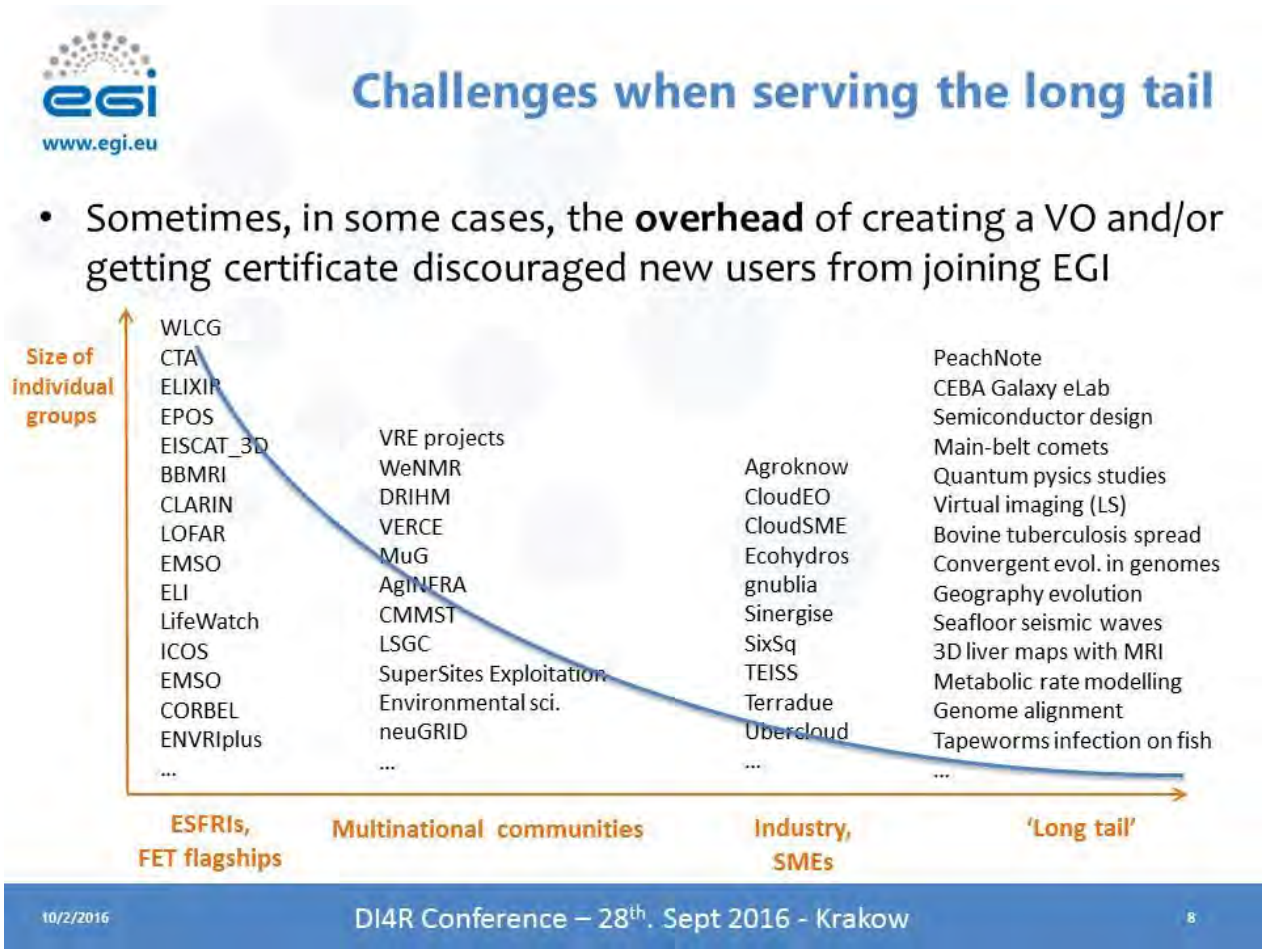
（出典）DI4R資料 Widmann, H. “EUDAT B2FIND :A Cross-Discipline Metadata Service and Discovery Portal”
http://www.digitalinfrastructures.eu/content/eudat-b2find-cross-discipline-metadata-service-and-discovery-portal

EGI: リソースと利用状況



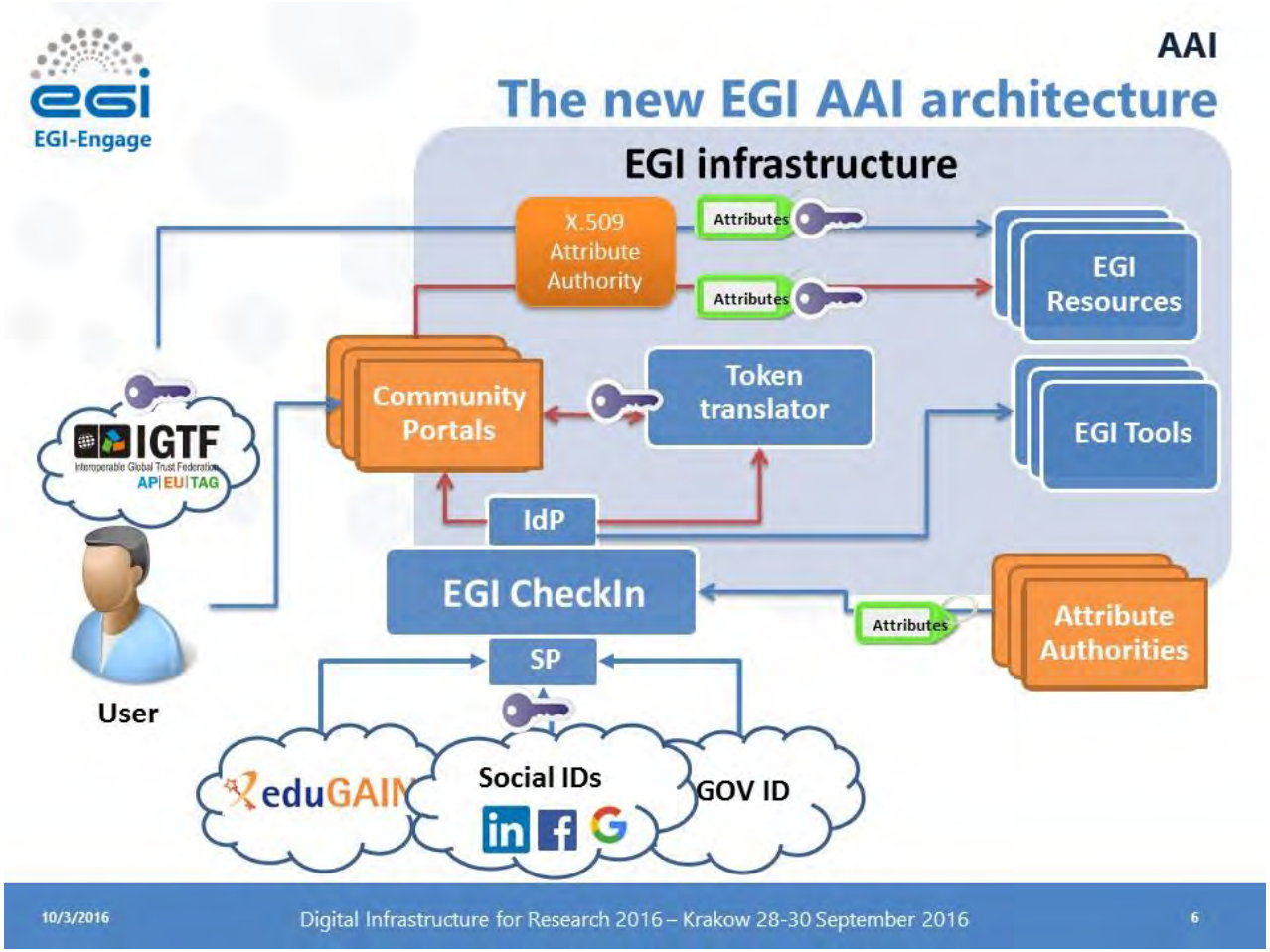
（出典）DI4R資料 La Rocca, G. “Serving the long tail”
http://www.digitalinfrastructures.eu/content/serving-long-tail

EGI: Cutting EdgeからLong Tailへ



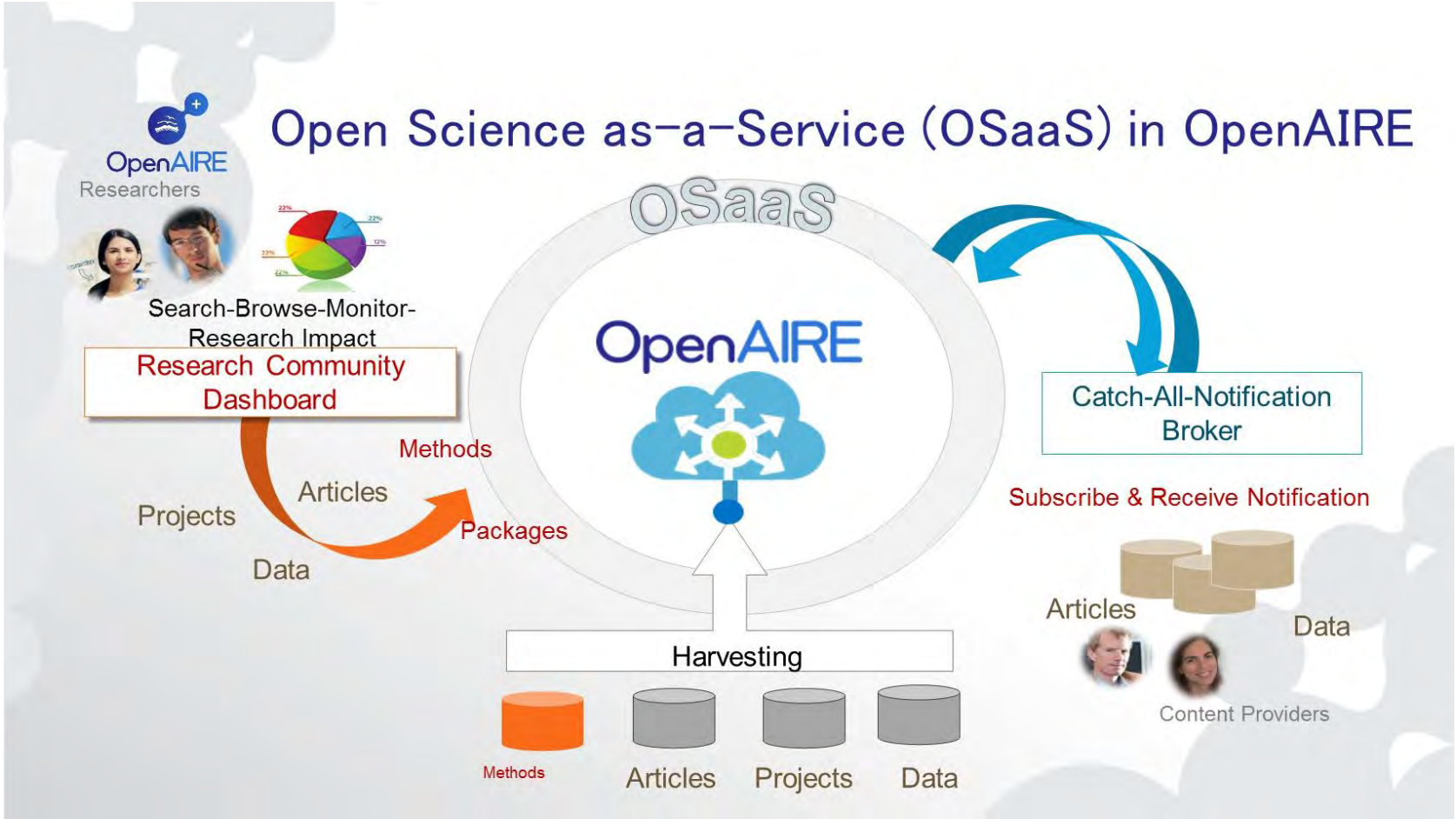
（出典）DI4R資料 La Rocca, G. “Serving the long tail”
http://www.digitalinfrastructures.eu/content/serving-long-tail

EGI: EGI CheckIn



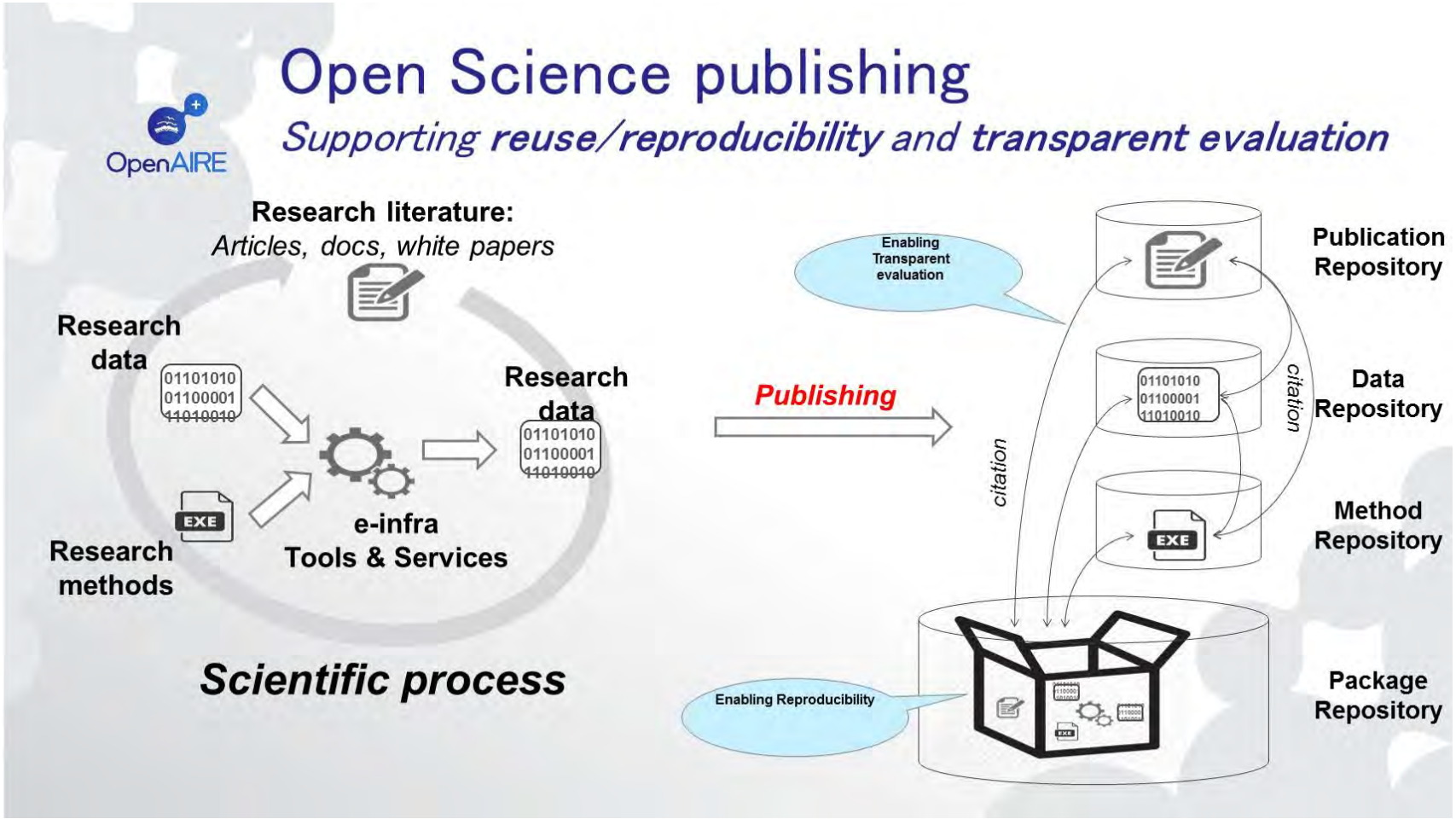
（出典）DI4R資料 Scardaci, D. “Solutions for federated services management EGI”
http://www.digitalinfrastructures.eu/content/egi-tools-and-solutions-federated-service-management

OpenAIRE: OSaaS



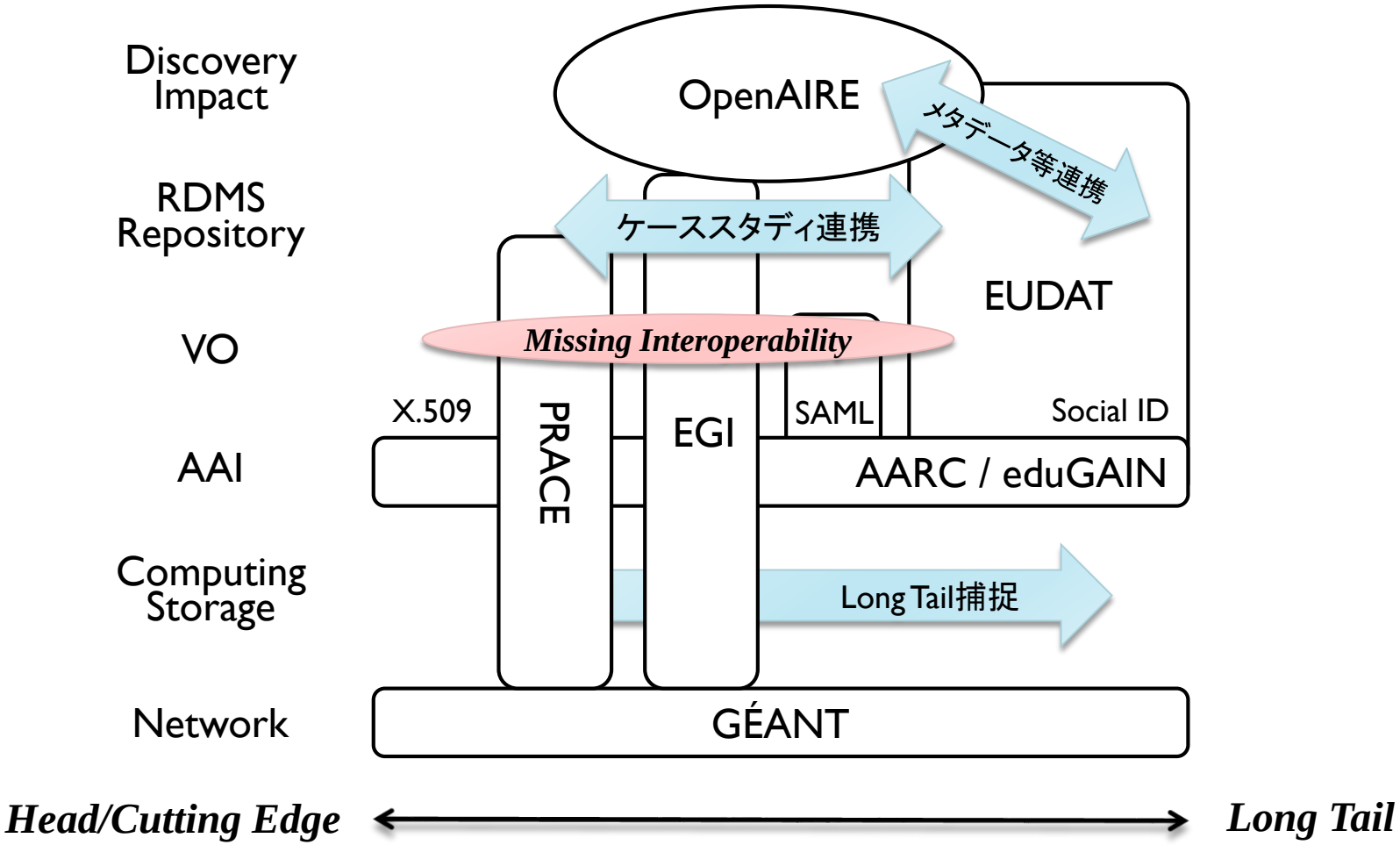
（出典）DI4R資料 Manghi, P. “OpenAIRE Infrastructure Services: technologies for Open Science”
http://www.digitalinfrastructures.eu/content/openaire-open-science-service

OpenAIRE: Open Scienceに向けて

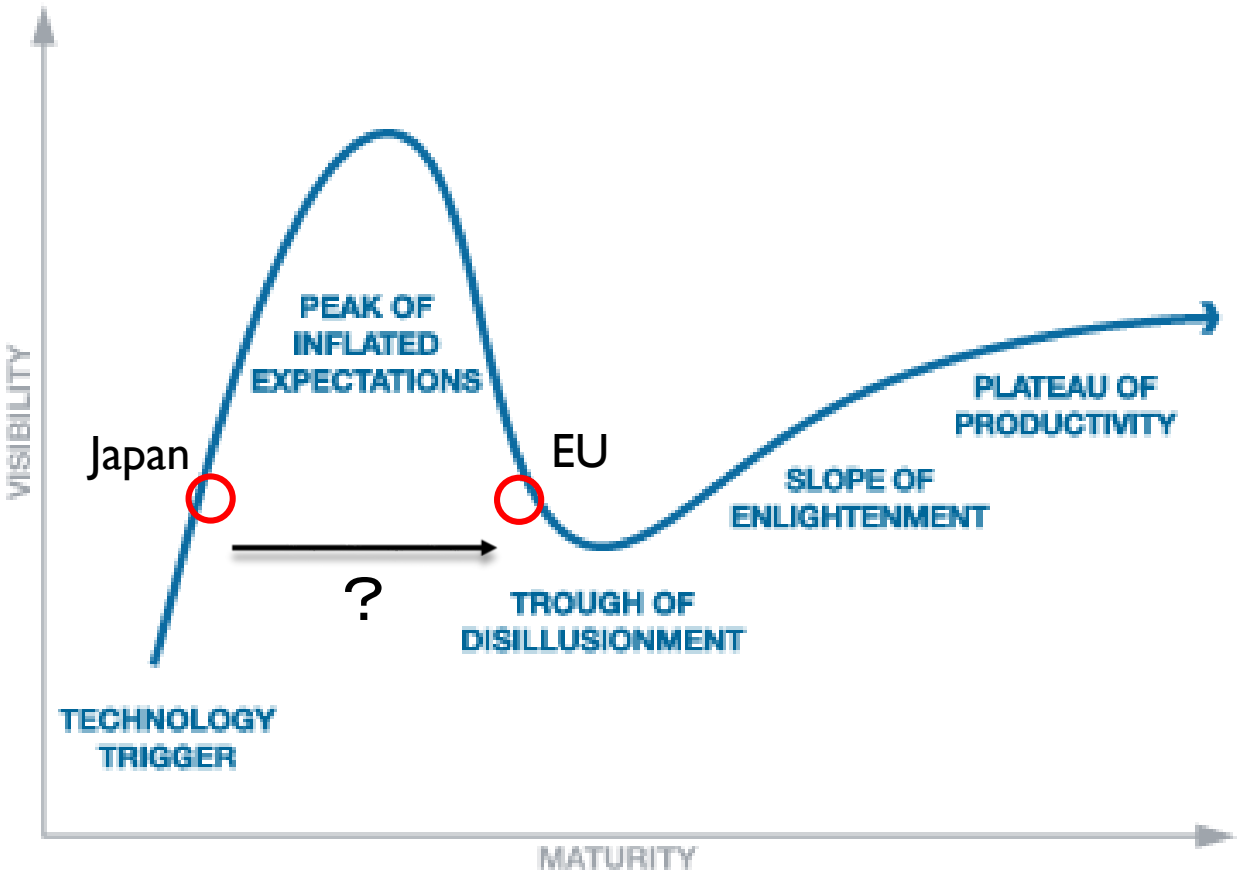


（出典）DI4R資料 Manghi, P. “OpenAIRE Infrastructure Services: technologies for Open Science”
http://www.digitalinfrastructures.eu/content/openaire-open-science-service

EOSCにおける統合の状況と課題

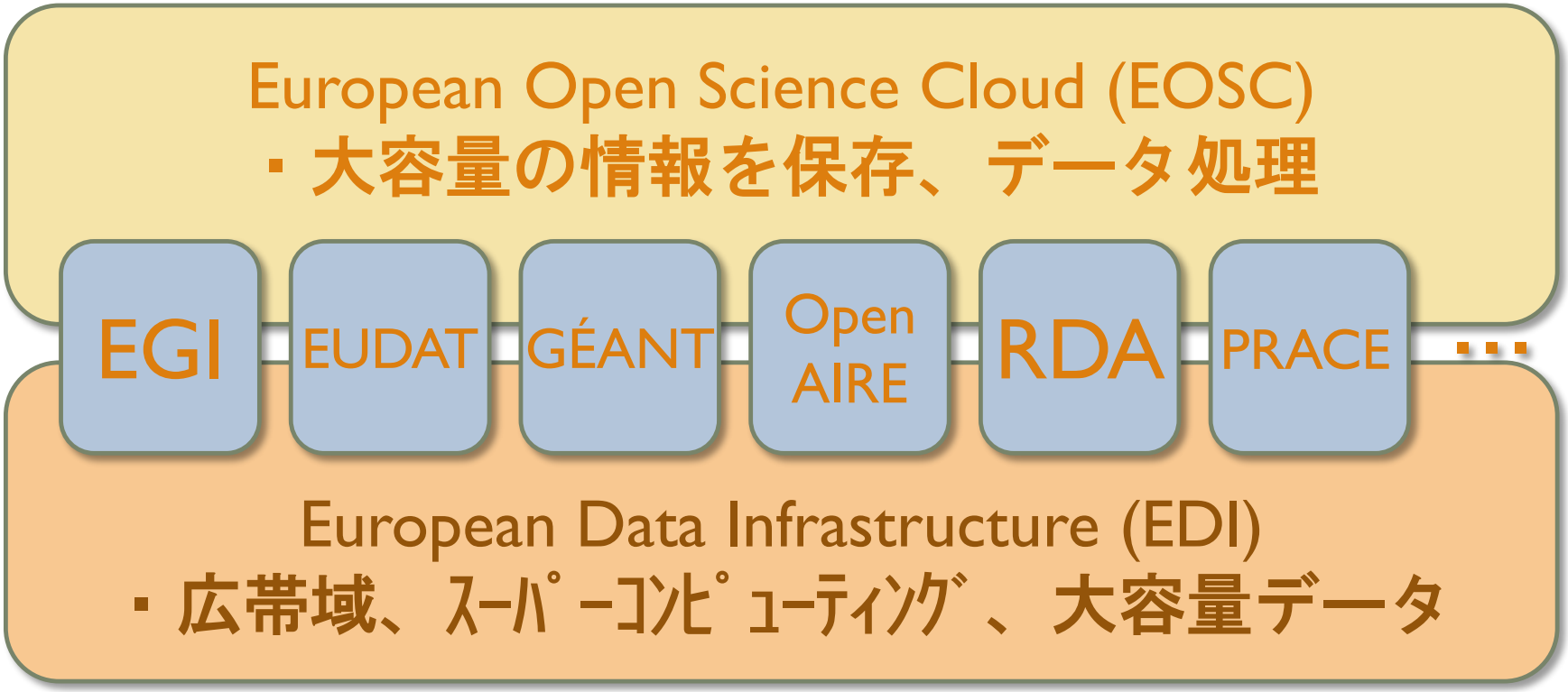


ギャップを克服できるか？



（出典）Gatner-HP “Gartner Hype Cycle”
http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

EUROPEAN CLOUD INITIATIVEの実現



- 既存のe-インフラを活用しつつ、間を埋め、データとコンピューティング・サービスを欧州の研究者とイノベーターに身近にする。

DI4Rのテーマ
—ロングテールへのアプローチ

- 背景
  - 既存のe-インフラは、ビッグサイエンスを中心に発展
  - しかし、あらゆる研究コミュニティにデジタル技術が裨益しないと、欧州の競争力は向上しない
- 課題
  - ビッグサイエンス以外の研究コミュニティは、VREを使いこなすための知識、スキルが欠落
  - それどころか、VREなど不要とばかりに、素通りする。

…そのようなコミュニティをどのように利用に促すか？

オープンサイエンスいろいろ
ーDI4R参加者にOSへのアプローチを聞いてみた

研究データと
論文がリンクし
RDMが円滑に
なること

もう
やっているから
特別なことは
不要

研究データの
透明性と
リユース

ビッグデータの
クラウド保存と
計算機能

EOSCで
足りない
機能は補完

結局のところ、
今までやっていることの延長線上にOSはある

オープンサイエンスへの示唆

- アプローチ
  - 既存のe-インフラを有効利用し、VREを実現
  - ビッグサイエンスからロングテールへ
  - データ管理と論文管理基盤を中心に実現
    - ✓ そうでないと、学術利用のウェブサービスをわざわざ構築する必要性が薄れる
- 課題
  - 日本は論文のOAや研究データ管理が十分に根付いていないため、利用拡大に難のある可能性
  - SME等学術界以外の利用はどうするか？
    - ✓ イノベーションを目的とするのなら・・・。