

オープンイノベーション、オープンサイエンスを支える永続的識別子の状況と未来への展望

蔵川圭

国立情報学研究所

<https://orcid.org/0000-0002-7031-1846>

発表の構成

- オープンイノベーションとオープンサイエンス
- オープンエコシステム
- 情報システムにおけるエンティティ
- エンティティと永続的識別子
- 永続的識別子の状況
- 未来への展望



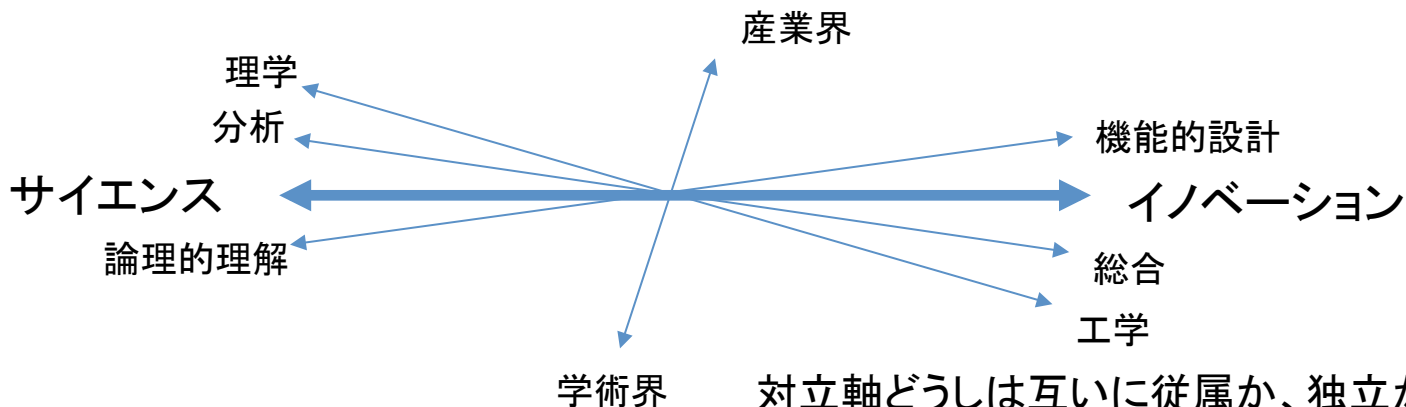
サイエンスとイノベーション

サイエンス

- 現象原理の論理的理解 (logical principle)
- 分析(analysis)
- 理学(science)
- 学術界(academia)

イノベーション

- 人工物の機能的設計 (functional design)
- 総合(synthesis)
- 工学(engineering)
- 産業界(industry)

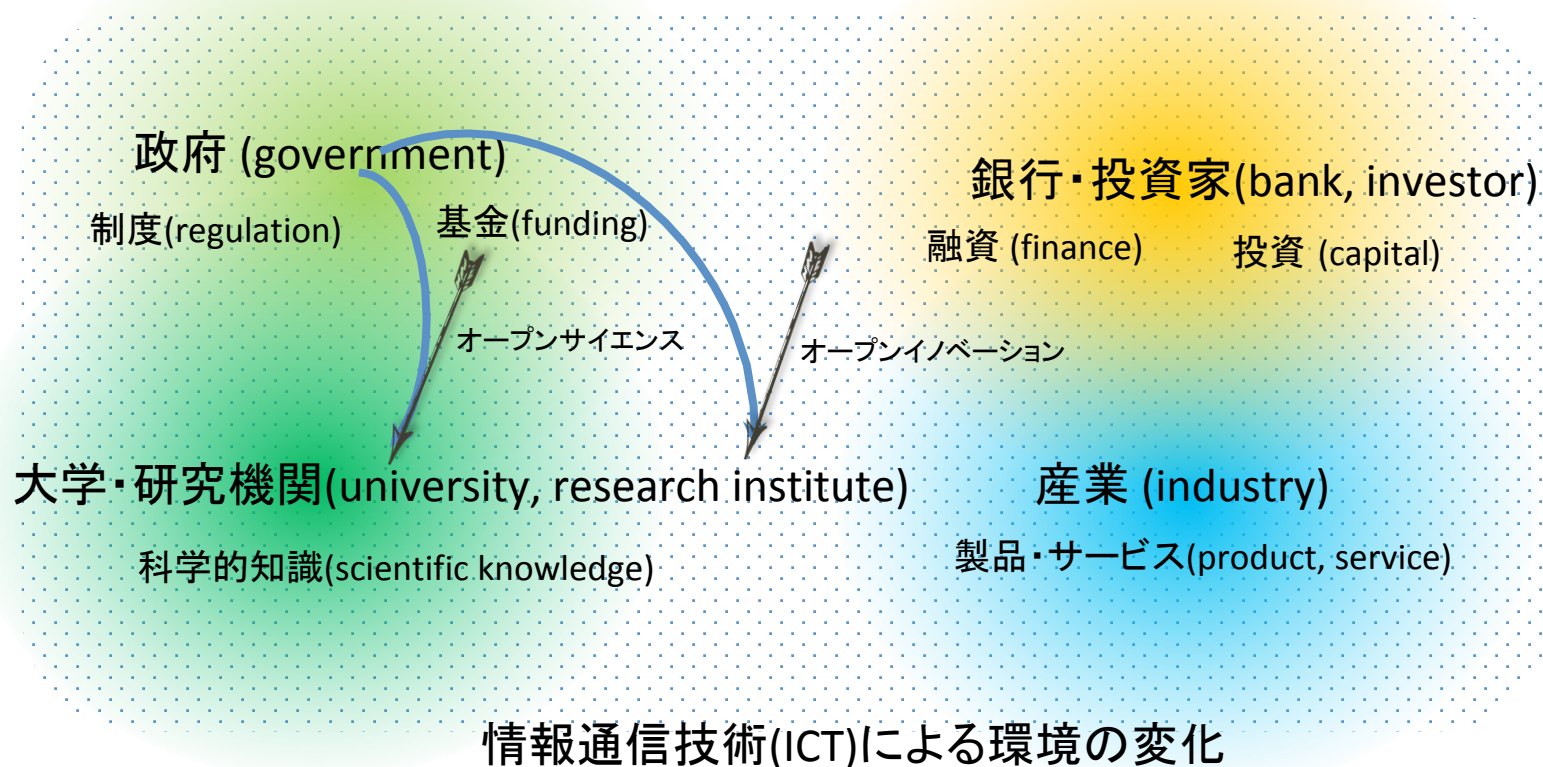


オープン、クローズド、シークレット

- 情報戦略の形
 - オープン(open)戦略
 - クローズド(closed)戦略
 - シークレット(secret)戦略
- 時代の文脈によって最適な戦略は変化する
 - 科学革命(17世紀)(Scientific Revolution)
 - 秘密主義、科学的知識は限られた専門家だけに共有される
 - 第二次世界大戦(1939-1945)(World War II)
 - 作戦司令の暗号化と解読
 - 新型兵器に関する軍事秘密
 - 冷戦時代(1945-1989)(Cold War)
 - 宇宙開発やインターネット
 - ポスト冷戦時代(1989-)(Post Cold War)
 - WWW、ソーシャルネットワーク
 - 雑誌論文の電子化とマネタイズ、トールアクセス(toll access)とオープンアクセス(open access)
 - 中国の台頭による、情報産業と科学的知識生産における勢力地図の変化
- オープン、クローズド、シークレットのバランスを取ることが大事
- デモクラシーとオープン戦略

学術に関する政策の2つの嚆矢

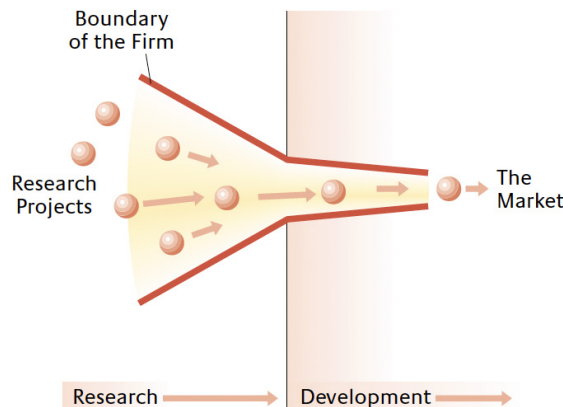
- オープンイノベーション
- オープンサイエンス



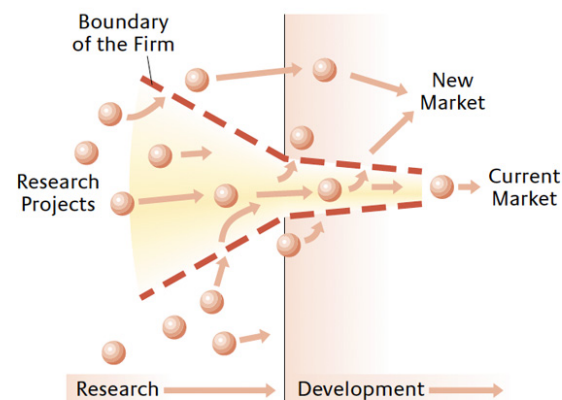
オープンイノベーション

- イノベーション研究(Innovation research)
 - 創造的破壊(creative destruction) [Joseph Schumpeter (1942)]
 - イノベーションとは、経済活動の中で生産手段や資源、労働力などをそれまでとは異なる仕方で新結合すること
 - 創造的破壊とは、経済発展というのは新たな効率的な方法が生み出されれば、それと同時に古い非効率的な方法は駆逐されていくという、その一連の新陳代謝
 - 破壊的イノベーション(disruptive innovation) [Clayton Christensen (1997)]
 - イノベーションには、従来製品の改良を進める「持続的イノベーション」と、従来製品の価値を破壊して全く新しい価値を生み出す「破壊的イノベーション」がある
 - オープンイノベーション(open innovation) [Henry Chesbrough (2003)]
 - 組織の枠を超えて迅速にイノベーションを誘発し最大経済効果を引き起こす情報時代のマネジメント手法

Closed innovation model



Open innovation model



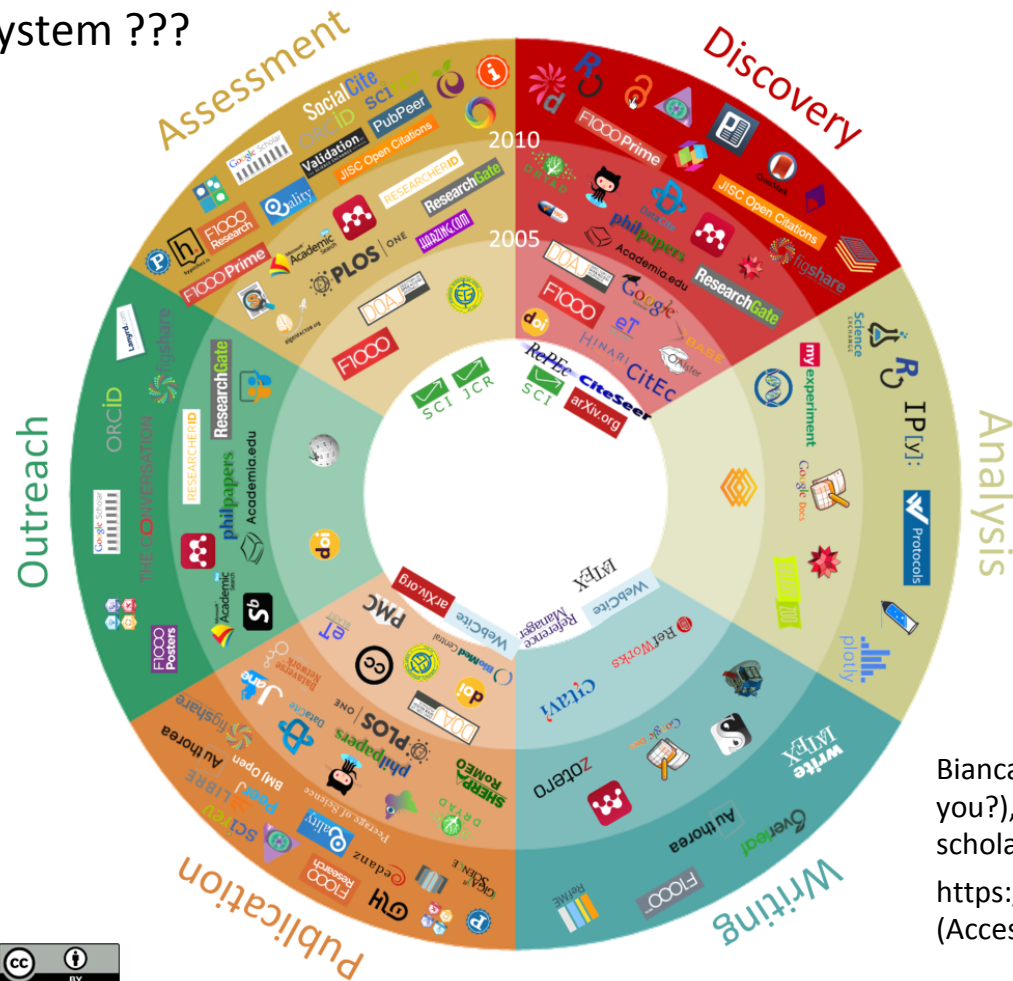
Henry Chesbrough (2003), The Era of Open Innovation, MIT Sloan Management Review, Spring 2003
<https://sloanreview.mit.edu/article/the-era-of-open-innovation/>

オープンサイエンス

- 17世紀の科学革命(Science Revolution)
 - アラビア・インド-ヨーロッパの数学と実験主義の融合[Paul A. David, 1994, 2008, 2014]
- 科学の規範 [Mertonian Norms(1942) and Ziman's extension(2000)]
 - Communism (公有主義)
 - Universalism (普遍主義)
 - Disinterestedness (利害の超越)
 - Originality (新規性)
 - Skepticism (懐疑主義)
- オープンサイエンス(Open Science)の歴史的起源 [Paul A. David, 1994, 2008, 2014, 2018]
 - サイエンスは中世ヨーロッパ封建制度における封臣から領主へのサービスから始まる
 - 現在は、オープンサイエンスとコマーシャルサイエンスが共存し、異なるシステムとして補完する
 - オープンサイエンスのシステムは物理学コミュニティに元からあるもの
 - 難解な数学のように科学の理解には情報的非対称性(informational asymmetry)があり、人物に知識があるかどうかはその人のレピュテーションによって初めて判断できる
 - 自分が何をしたかを示すには知識をオープンにし、レピュテーションを得るにはその最初でなければならない
 - しかし、ニュートンでさえ不十分な知識を公開することは危険であるためよくないと考えていた
- 'Open Science' : ネットワーク化された科学 (networked science) [Michael Nielsen, 2011]
 - 集合知 (collective knowledge)の活用
 - 知識やデータのオープンアクセス
 - シチズンサイエンス
 - オープンサイエンス革命(open science revolution)

オープン・エコシステム

Open Science Ecosystem ???



Bianca Kramer & Jeroen Bosman (and you?), 400+ Tools and innovations in scholarly communication

<https://101innovations.wordpress.com/>
(Accessed: 2018-06-15)

エコシステム連携

- なぜエコシステム連携？
 - システムレベルの新しい価値の創出
 - マーケットニーズへの迅速な対応
 - オープンイノベーションの実現
- 技術的特徴
 - サブシステムどうしのサービス連携、データ連携
 - データの蓄積、統合（ビッグデータ処理）
 - 分析（アナリティクス）

エンティティ

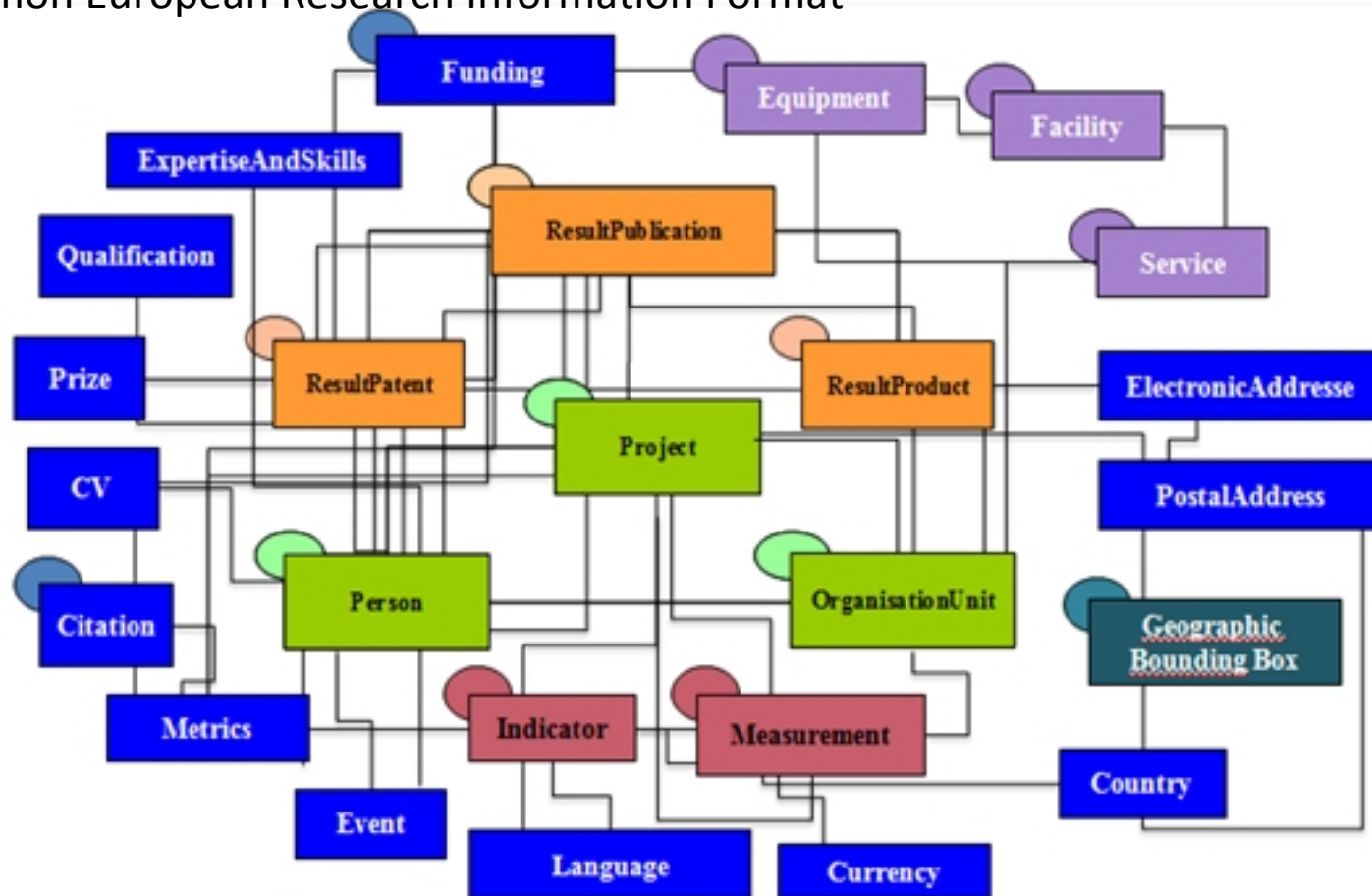
- エンティティとは
 - サービスやデータとして何を扱うか
 - 情報の概念としての対象
- エンティティの識別
 - 名前
 - ラベル
 - 識別子
- 識別されたエンティティの関係構造
 - 辞書
 - グラフ
 - 概念オントロジー
- エンティティグラフと概念階層
 - データ (data)
 - データ構造
 - 情報 (information)
 - ER図
 - 知識 (knowledge)
 - ナレッジグラフ

エンティティの種類

- エンティティの対象
 - 我々が思考可能な存在そのもの
 - 辞書項目のように膨大
- 科学技術、学術の管理におけるエンティティ
 - EuroCRISによるCERIF(the Common European Research Information Format)が参考になる
 - Current Research Information Systems
 - <https://www.eurocris.org>

CERIF

the Common European Research Information Format



<https://www.eurocris.org/cerif/main-features-cerif>

エンティティ識別の方法

- 識別子の付与 (identifier)
 - エンティティクラスを定義して識別子空間を設計し、エンティティ要素に識別子を付与
 - 離散的モデル(discrete model)
- 名寄せ (name disambiguation)
 - エンティティ依存の距離空間を定義し、エンティティ要素を分類
 - 階層的クラスタリング
 - クラスラベリング
 - 確率的モデル(probabilistic model)

識別子の永続性

- 識別子の永続性の議論
 - Kahn/Wilensky Architecture (1995), doi:cnri.dlib/tn95-01
 - A Framework for Distributed Digital Object Services
 - ロケーションレイヤーと識別子レイヤーを分ける
 - Handle System (CNRI)
 - DOI System (International DOI Foundation)
 - » CrossRef
 - » DataCite
 - » JaLC
 - ブログにおけるパーマリンク (2000ごろ)
 - トラックバックによる記事相互参照のメカニズム
 - Web2.0 [Darcy DiNucci, 1999]
 - WWW Architecture [W3C, 2004, <http://www.w3.org/TR/webarch/>]
 - Cool URIs don't change! [Tim Berners-Lee, 1998]
 - URLとURI, URN
 - HTTP/HTMLからSemantic Webへ [W3C, 2008, <https://www.w3.org/TR/cooluris/>]

永続的識別子(PID)名前空間のタイプ

- 汎用
 - World Wide Web上
 - URI [RFC1630, RFC3986] (operated by web server manager)
 - URN [RFC1737, RFC8141] (operated by the Internet Assigned Numbers Authority (IANA))
 - PURL (operated by OCLC, now the Internet Archive)
 - ARK (operated by California Digital Library)
 - デジタル空間上
 - Handle [RFC3650, RFC3651, RFC3652] (operated by the DONA)
 - DOI (operated by the International DOI Foundation)
- 個別エンティティ用
 - ISBN
 - ORCID
 - IGSN (International Geo Sample Number)
 - LSID (Life Science identifiers)
 - Chemical Entities of Biological Interest (ChEBI)
 - Perma.cc (Archived Web page especially for law studies and judicial opinions)
 - etc

PIDリゾルバー

- Handleリゾルバー
 - Handle.Net
 - Corporation for National Research Initiatives (CNRI)
 - Now, the DONA foundation in Switzerland
 - DOI system Proxy Server
 - the International DOI Foundation
 - The DOI system uses the Handle System® to manage digital objects.
- Prefixed PIDリゾルバー (メタリゾルバー)
 - N2T Resolver
 - California Digital Library
 - Identifiers.org
 - The European Bioinformatics Institute (EMBL-EBI)
 - Identifiers.org share a common prefix registry with N2T.net, based at the California Digital Library, enabling users to resolve Compact Identifiers using the Identifiers.org or N2T.net meta-resolver.

名前空間上に登録されるエンティティ

PIDのエンティティタイプ

研究リソース

研究者

組織

グラント

デジタルオブジェクト、研究データ、研究サンプル、研究機器?、コンセプト、語彙、分類

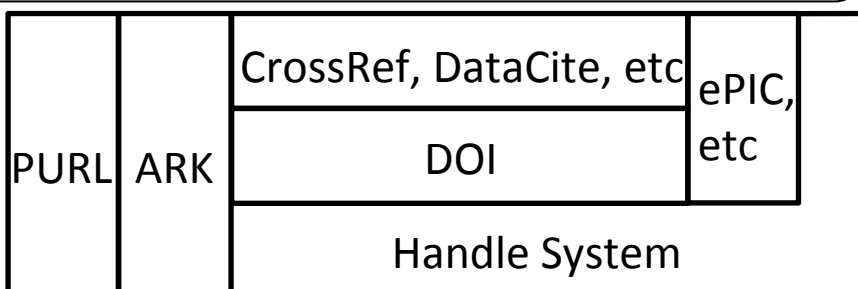
OrgID?、プロジェクト?

Meta-resolver / Handle, DOI, ARK, PURL resolver

URI / URN

ISBN,
LSID,
ChEBI,
Perma.cc,
etc

Publisher articles / figures,
Data citation,
IGSN, etc



eduPersonOrgOrcid

ORCID

eduPersonOrgDN

ISNI?
GRID?
Ringgold?

CrossRef Funder?

Federated
Identity
Management

GRID

Global Research Identifier Database

Cataloging the world's research organisations

[DOWNLOAD THE FREE DATABASE](#)

Operated by Digital Science & Research Solutions Ltd
at <https://www.grid.ac>

MAKE SENSE OF YOUR INSTITUTIONAL DATA

Capture data accurately



Ensure robust reporting



Disambiguate your data





Identifiers.org

Search

[Advanced search](#)[Home](#)[Documentation](#)[Services](#)[About](#)[Feedback](#)

Identifiers.org is an established resolving system that enables the referencing of data for the scientific community, with a current focus on the Life Sciences domain. It handles persistent identifiers in the form of URIs and CURIEs. This allows the referencing of data in both a location-independent and resource-dependent manner. The provision of resolvable identifiers (URLs) fits well with the Semantic Web vision, and the Linked Data initiative.

Resolve Compact Identifiers (prefix:identifier), eg: CHEBI:36927



Registry



Request prefix



Web Services



Download

Data records

Collections	647
Resources	805
Last updated	2018-05-21

Meta-resolvers

Identifiers.org share a common prefix registry with N2T.net, based at the California Digital Library, enabling users to resolve Compact Identifiers using the Identifiers.org or N2T.net meta-resolver.

Connect with us

[Twitter](#)[GitHub](#)

Tweets by @IdentifiersOrg

[identifiers.org](#) Retweeted**Scientific Data**
@ScientificData

We are now using the services provided by [identifiers.org](#) & [n2t.net](#) to improve our data citations. Read more about this change and the importance of good [#datacitation](#): [nature.com/articles/sdata...](#)



On the road to robust da...
On the road to robust data...
[nature.com](#)



May 9, 2018

[identifiers.org](#) Retweeted**ELIXIR Europe**
@ELIXIREurope

The new compact identifiers to cite [#lifescience](#) data developed by [@IdentifiersOrg](#) & [n2t.net](#) have been adopted by [@ScientificData](#): [goo.gl/AJP5yd](#) Also thanks to ELIXIR: [goo.gl/QCnuDw](#) [@CalDigLib](#) [@OfficialUoM](#) [@emblabi](#) [@UVADSI](#)

Article | [OPEN](#)

Uniform resolution of compact identifiers for biomedical data

Sanja M. Winkler, Nick Juty, John Krumm, Greg Jander, Julie A. Hetherington, Niall Beal, Rafael Jimenez, Jeffrey S. Gresh, Henning Hermjakob, Margarete E. Martens & Tim Clark



Identifiers.org is part of the ELIXIR infrastructure

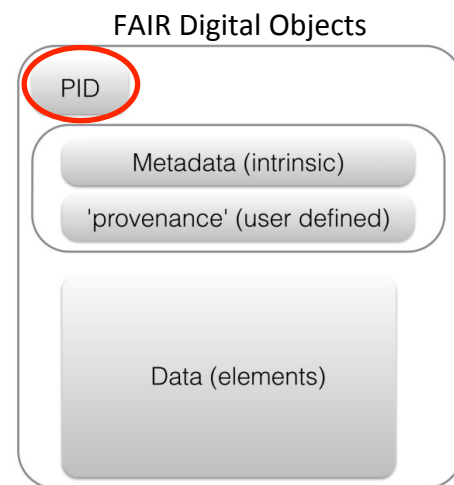
[Learn more](#)

id	name	pattern
MIR:00000002	ChEBI	^CHEBI:\d+\$
MIR:00000003	Ensembl	^((ENS[FPTG]\\d{11}(\\.\\.\\d+)?) (FB\\w{2}\\d{7}) (Y[A-Z]{2}\\d{3}[a-zA-Z](\\-[A-Z])?) ([A-Z_a-z0-9]+(\\.\\.)?(t)?(\\d+)?([a-z])?))\$
MIR:00000004	Enzyme Nomenclature	^\\d+\\.\\.\\.\\- \\d+\\.\\.\\d+\\.\\.\\.\\- \\d+\\.\\.\\d+\\.\\.\\d+\\.\\.\\- \\d+\\.\\.\\d+\\.\\.\\d+\\.\\.\\d+\\.\\(n)?\\d+\$
MIR:00000005	UniProt Knowledgebase	^([A-N,R-Z][0-9]([A-Z][A-Z, 0-9][A-Z, 0-9][0-9]){1,2}) ([O,P,Q][0-9][A-Z, 0-9][A-Z, 0-9][A-Z, 0-9][0-9])(\\.\\.\\d+)?\$
MIR:00000006	Taxonomy	^\\d+\$
MIR:00000007	BioModels Database	^((BIOMD MODEL)\\d{10}) (BMID\\d{12})\$
MIR:00000008	MIRIAM Registry collection	^MIR:000\\d{5}\$
MIR:00000009	ICD	^[A-Z]\\d+(\\.\\.\\-\\d+)?\$
MIR:00000010	IntAct	^EBI\\-[0-9]+\$
MIR:00000011	InterPro	^IPR\\d{6}\$
MIR:00000012	KEGG Pathway	^\\w{2,4}\\d{5}\$
MIR:00000013	KEGG Compound	^C\\d+\$
MIR:00000014	KEGG Reaction	^R\\d+\$
MIR:00000015	PubMed	^\\d+\$
MIR:00000016	OMIM	^[*#+%^]?\\d{6}\$
MIR:00000017	PIRSF	^PIRSF\\d{6}\$
MIR:00000018	Reactome	(^R-[A-Z]{3}-\\d+(-\\d+)?(\\.\\.\\d+)?\$) (^REACT_\\d+(\\.\\.\\d+)?\$)
MIR:00000019	DOI	^(doi\\:)?\\d{2}\\\\.\\d{4}\\.*\$
MIR:00000020	Protein Data Bank	^[0-9][A-Za-z0-9]{3}\$
MIR:00000022	Gene Ontology	^GO:\\d{7}\$
MIR:00000023	SGD	^((S\\d+\$) (Y[A-Z]{2}\\d{3}[a-zA-Z](\\-[A-Z])?))\$
MIR:00000024	Systems Biology Ontology	^SBO:\\d{7}\$
MIR:00000025	KEGG Drug	^D\\d+\$
MIR:00000026	KEGG Glycan	^G\\d+\$
MIR:00000027	WormBase	^WB[A-Z][a-z]+\\d+\$
MIR:00000028	Pfam	^PF\\d{5}\$
MIR:00000029	Nucleotide Sequence Database	^([A-Z]\\d{5} [A-Z]{2}\\d{6} [A-Z]{4}\\d{8} [A-J][A-Z]{2}\\d{5})(\\.\\.\\d+)?\$
MIR:00000030	FlyBase	^FB\\w{2}\\d{7}\$
MIR:00000031	Wormpep	^CE\\d{5}\$
MIR:00000032	PROSITE	^PS\\d{5}\$
MIR:00000033	PubChem-substance	^\\d+\$
MIR:00000034	PubChem-compound	^\\d+\$
MIR:00000035	arXiv	^(\\w+(\\-\\w+)?(\\ \\w+)?/)?\\d{4}\\ Z\\(\\ \\d{4}\\(v\\d+)?\\)?\$

<http://identifiers.org/rest/collections/>
, accessed at 2018-06-11

データをFAIRに

- FAIR原則[Winkinson MD et.al. 2016, NDBC研究チーム, 2018]
 - To be Findable: (見つけられるために)
 - F1. (メタ)データが、グローバルに一意で永続的な識別子(ID)を有すること。
 - F2. データがメタデータによって十分に記述されていること。
 - F3. (メタ)データが検索可能なリソースとして、登録もしくはインデックス化されていること。
 - F4. メタデータが、データの識別子(ID)を明記していること。
 - To be Accessible: (アクセスできるために)
 - A1. 標準化された通信プロトコルを使って、(メタ)データを識別子(ID)により入手できること。
 - A1.1 そのプロトコルは公開されており、無料で、実装に制限が無いこと。
 - A1.2 そのプロトコルは必要な場合は、認証や権限付与の方法を提供できること。
 - A2. データが利用不可能となったとしても、メタデータにはアクセスできること。
 - To be Interoperable: (相互運用できるために)
 - I1. (メタ)データの知識表現のため、形式が定まっていて、到達可能であり、共有されていて、広く適用可能な記述言語を使うこと。
 - I2. (メタ)データがFAIR原則に従う語彙を使っていること。
 - I3. (メタ)データは、他の(メタ)データへの特定可能な参照情報を含んでいること。
 - To be Re-usable: (再利用できるために)
 - R1. メタ(データ)が、正確な関連属性を豊富に持つこと。
 - R1.1 (メタ)データが、明確でアクセス可能なデータ利用ライセンスと共に公開されていること。
 - R1.2 (メタ)データが、その来歴と繋がっていること。
 - R1.3 (メタ)データが、分野ごとのコミュニティの標準を満たすこと。



[FORCE11, 2014]

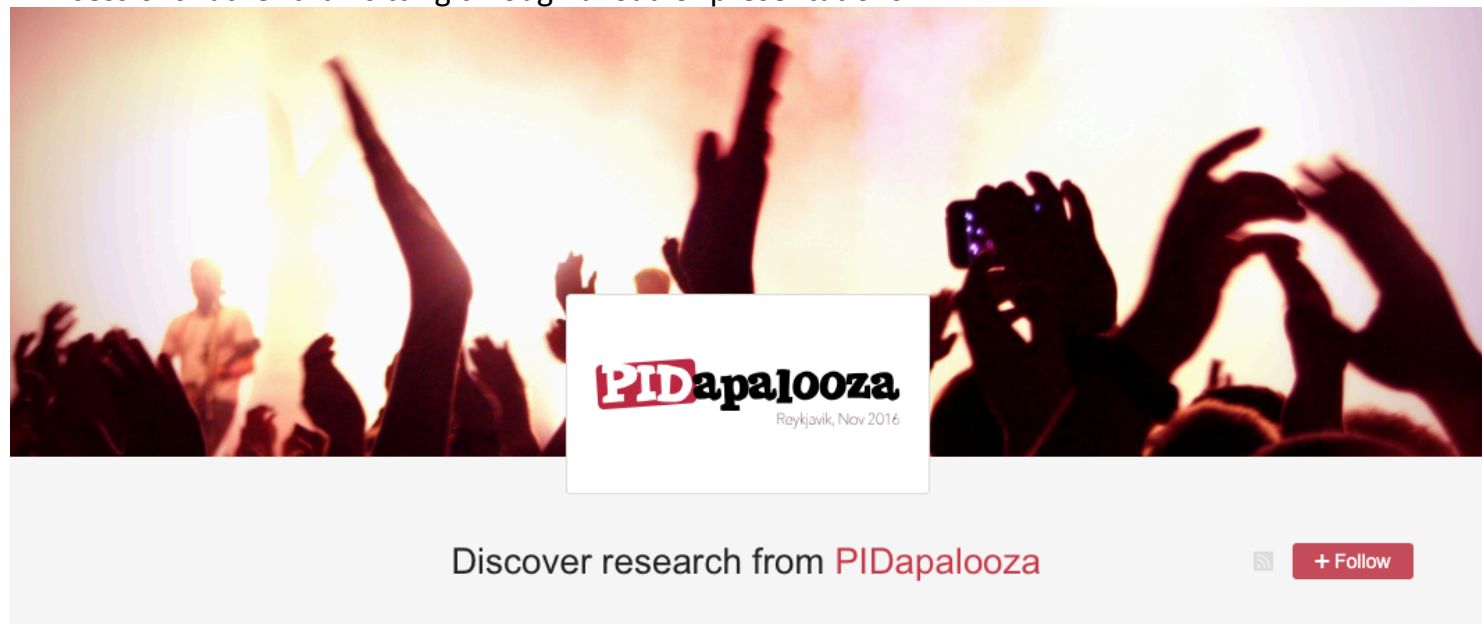
FORCE11, GUIDING PRINCIPLES FOR FINDABLE, ACCESSIBLE, INTEROPERABLE AND RE-USABLE DATA PUBLISHING VERSION B1.0, 2014, <https://www.force11.org/fairprinciples>

Wilkinson MD et. al.: The FAIR Guiding Principles for scientific data management and stewardship. Sci Data., doi:10.1038/sdata.2016.18 (2016)

NDBC研究チーム、データ共有の基準としてのFAIR原則、2018-04-19、DOI:10.18908/a.2018041901

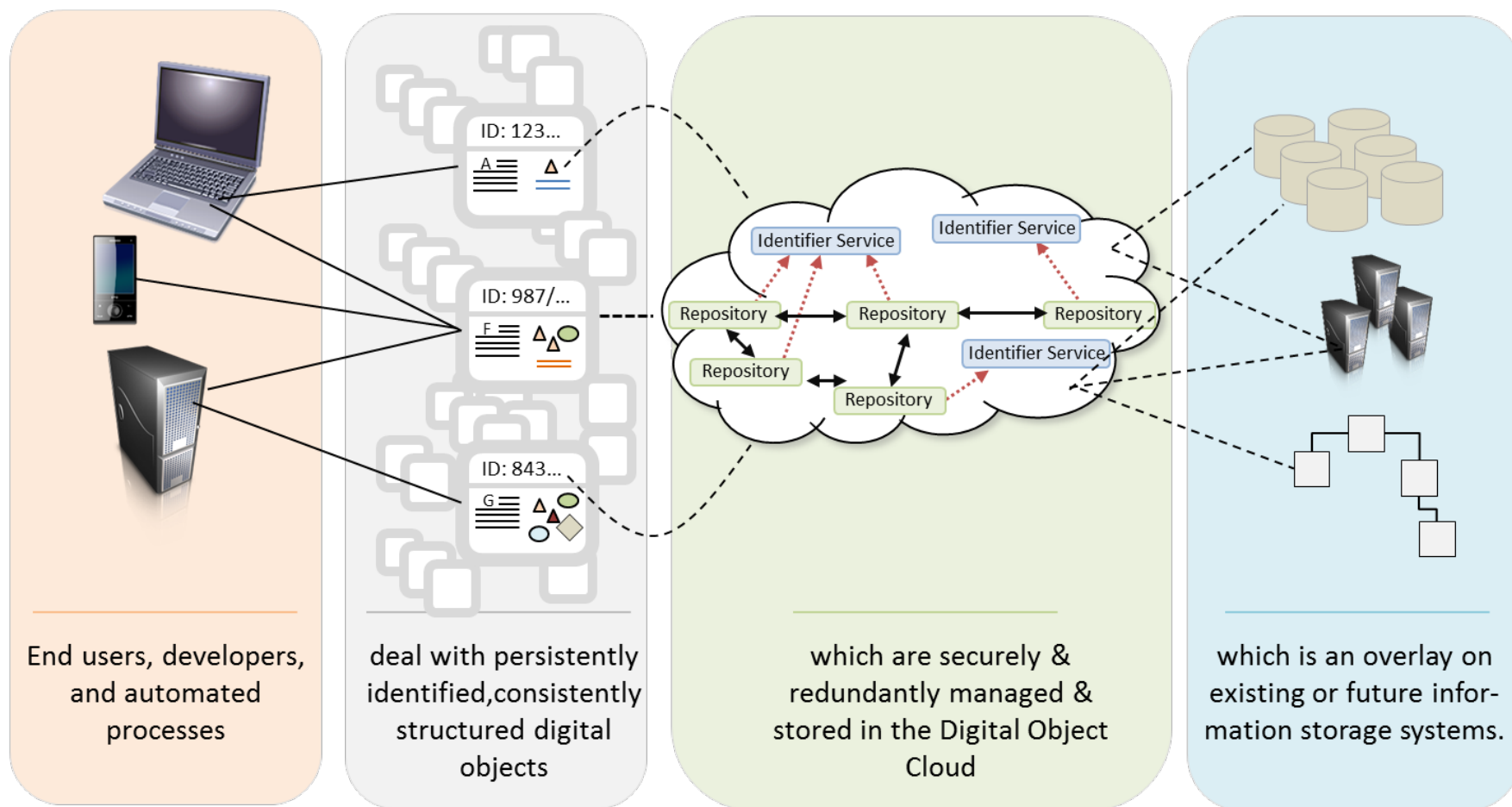
PIDイベント

- PIDapalooza (2016, 2018)
- Christopher Brown, A Festival for Persistent Identifiers, Research Data and related topics, 2018-02-06, at <https://researchdata.jiscinvolve.org/wp/2018/02/06/pidapalooza18/>, accessed at 2018-06-11
 - “Open identifiers deserve their own festival” say the organisers of PIDapalooza and if you have any interest in Persistent Identifiers (PIDs) then this is the event for you. With a unique style roughly modelled on a music festival (nail painting and tattoos were optional), the focus is on discussions and questions in short, parallel sessions rather than sitting through a load of presentations.



56 presentation slides are available at <https://pidapalooza.figshare.com>

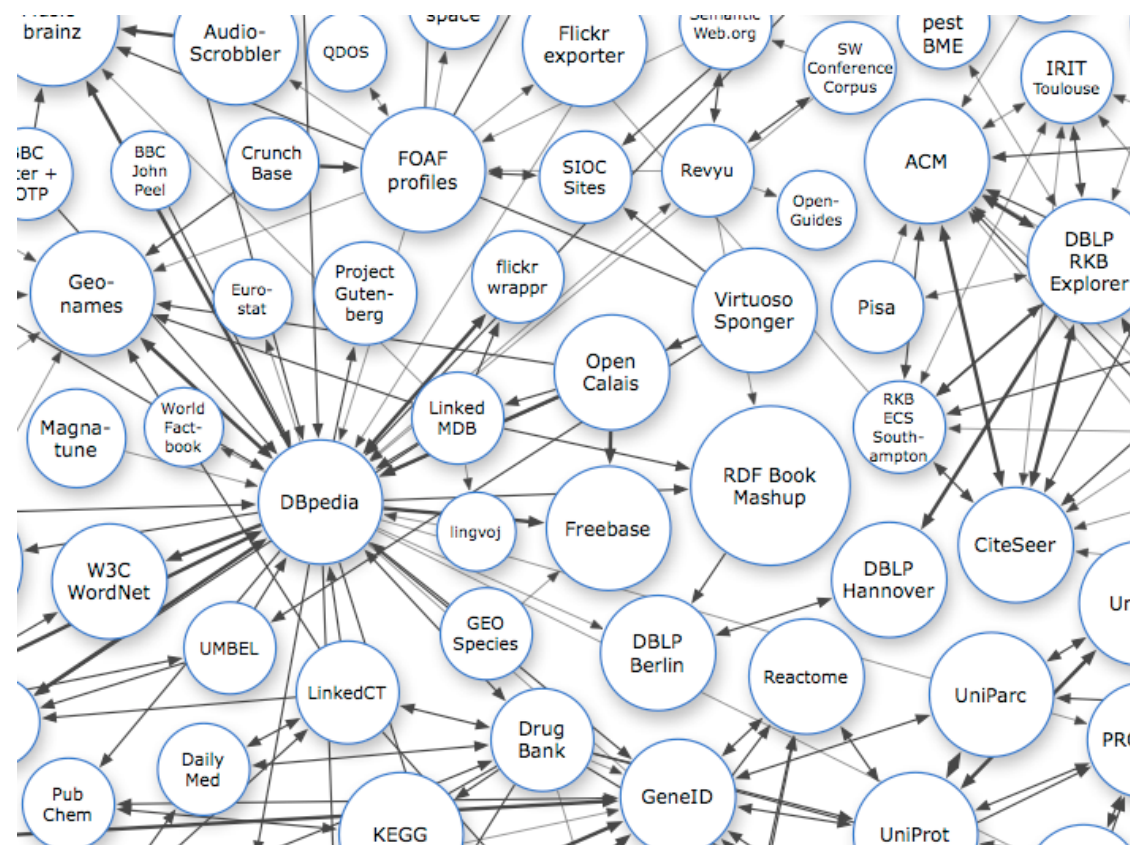
デジタルオブジェクトクラウド



Larry Lannom, Peter Wittenburg, Global Digital Object Cloud (DOC) - A Guiding Vision, 11 September 2016,
<http://hdl.handle.net/11304/a8877a1a-9010-428f-b2ce-5863cec4aff3>

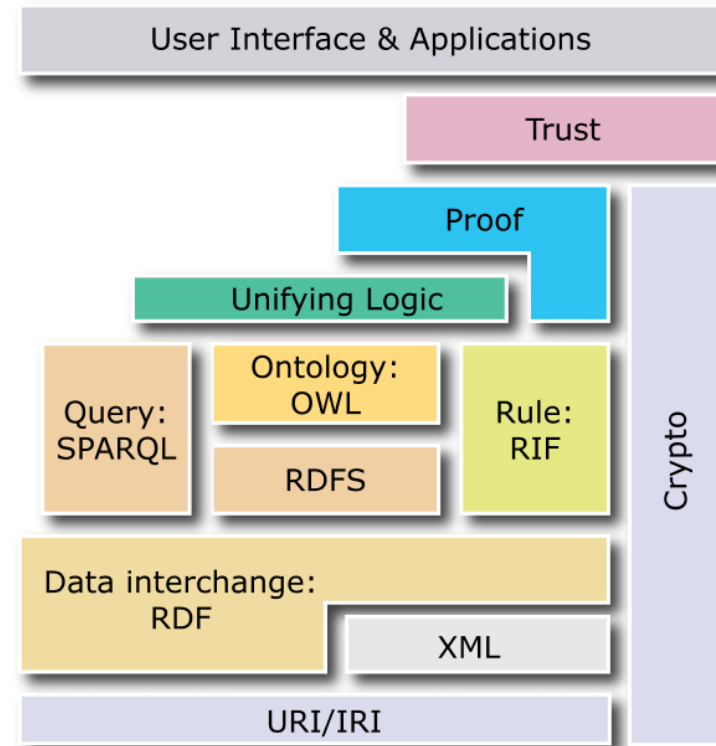
リンクトデータ

Linked Data - Connect Distributed Data across the Web



<http://linkeddata.org>

Semantic technology layer cake



<https://www.w3.org/2007/03/layerCake.png>

豊かな情報社会を目指す法規制

- 情報利活用の推進と法規制
 - Lawrence Lessig の“Code and Other Laws of Cyberspace” (1999)や、社会実装としてのCreative Commons Licenseはさきがけ的存在
- 情報社会への期待と法規制
 - 2018年の時価総額ランクトップ10に、Apple, Amazon, Microsoft, Alphabet (Google), Facebook, Tencent, AlibabaのIT7社
 - 改正個人情報保護法
 - 匿名加工情報の活用、再識別の禁止
 - 2015年9月改正、2017年5月全面施行
 - http://www.meti.go.jp/policy/it_policy/privacy/
 - EU一般データ保護規則GDPR (The EU General Data Protection Regulation)
 - EU域内における個人情報保護規則の共通化
 - 個人情報の範囲と取得、管理ルールが強化
 - 罰則の強化
 - https://ec.europa.eu/info/law/law-topic/data-protection_en

未来への展望 (1/2)

- ネットワーク化、ボーダレス化、オープン化する社会
 - ネットワーク化された科学・技術、学術
 - 組織を超えた社会的イノベーション
- ユニバーサルな情報アクセス(Universal Information Access)を目指して
 - 情報の組織化としてのデジタルオブジェクトクラウド
 - 知識表現としてのリンクトデータ
 - 労働集約型データベースの超分散構築
 - 多様なエンティティを共通の情報基盤技術の上に表現
 - 永続的識別子を交点とするSemantic Webとデジタルライブラリ技術のWeb上の混合

未来への展望 (2/2)

- スマート社会の実現に向けて
 - 科学・技術、学術のデータ、情報、知識クラウド
 - セクターを超えるオープンエコシステム
 - 組み合わせ自由の知的情報処理への挑戦
 - コード化(制度化)される情報社会
 - 社会規範の再確認と情報活用の強化
 - 基本的人権 (human rights)
 - 科学の規範 (norms of science)
 - 学問の自由 (academic freedom)
 - セクターの役割
 - » 産業界
 - » 官界
 - » 学界
 - 資本主義、社会主義、情報主義
 - 情報戦略のオープン、クローズド、シークレット