

# ③ 地球惑星科学の視点でみるオープンサイエンス—研究データの取り扱いを振り返る—

村山泰啓 | 情報通信研究機構

## オープンサイエンスにおけるデータ

現在、国内でも議論が活発になってきている「オープンサイエンス」において、研究データの取り扱い、いわば「新しいデータマネジメント」は大きなテーマの1つである。国際政治の場では、2013年G8科学技術大臣会合で研究データオープン化が合意されて以降、日本国内では内閣府が中心となって国際動向にかかわるオープンサイエンスの推進を議論してきた。筆者は2013年頃からこれにかかわり、2016年、2017年のG7における、より具体的な国際科学ポリシーにおける、オープンサイエンスの議論に参加する機会を得た。

図-1は、この議論において科学データの取り扱いの基本的考えを示したスライドで、科学・技術等にとってデータは科学と社会の相互信頼のための重要な要素であり、また人類が共有すべき情報資産である、といった、欧米の識者の間ではここ10年ほ

### Data

- Data issues
  - Mutual trustworthiness of Science and Society
  - Information asset for the human society
  - Fuel to drive/accelerate science & technology
    - Data Driven Innovation (OECD, etc.)
  - Data as a "first class" research output/resource
  - What is the best practice for both Science and Society?

(Image courtesy of Wikimedia Commons.)



A. Einstein, B. Rodolfsky, and N. Rosen(1935)

(By Yuri Samoilov, Creative Commons Attribution 3.0)

図-1 国際的な科学研究データの取り扱いに関する基本的考え方のまとめ例 (村山, 2016).

どのあいだに科学政策やアカデミーの視点で議論されてきた共通理念を書いている。2016年のG7科学技術大臣会合で話題提供講演をしたときのものであるが、G7各国政府関係者にとってはなじみのある論点だったのだと思われる。その後も内閣府およびG7科学技術大臣会合の枠組みにおいて、これを受けた議論に協力させていただいている。

なお「科学」「科学・技術等」と書いているものは、自然科学やいわゆる理工系に閉じる話ではなく、生命科学、社会科学、人文学などのほぼすべての科学・学術分野を想定している。「研究データ」とは、そうした広範な分野の、専門的で知的生産性（可能性）の高い情報アセット（資産）全般と考えている。

## 科学とデータ

図-1の下半分に Albert Einstein 博士の写真を掲載している。これは、近年発展してきた量子暗号通信というテクノロジー分野の論文において、重要な理論的基礎を与えていて数多く引用されている論文が、この Einstein らの1935年の論文であるということを示すためのものだ。若手研究者も、この80年以上前の論文のPDFファイルは、インターネット経由で容易に手に入れられる。現代のテクノロジー、イノベーションが、証拠と合理性に基づいた論理構築という近代科学の原理をツールとして実現して活用されている今日、研究の根拠や元情報が参照できることによって、最新テクノロジーがいつも科学という巨大知的構築物の柱や構造材に支えら

れ続けているといってもよいだろう。

2016年のG7科学技術大臣会合で述べた言葉を思い出せる限り日本語にして書くと以下のようなものである：現代のほとんどのイノベーションとその基礎は、過去に紙に印刷された論文の蓄積によって可能となりましたが、いまある高度なテクノロジーの検証と発展においてデータは不可欠です。これから80年後、私たちの孫やその子孫がイノベーションを可能としていくために今後80年間どうやってデータを残していくのかは、大変重要な問題ではないでしょうか——こう申し上げた。

## 地球科学のデータ

オープンサイエンス、という新しい概念（はやり言葉なのか新しい科学の枠組みになるか、まだ分からないが）が議論される以前から、当然、研究データの取り扱いにはさまざまな議論や取り組みがあった。

私たち人類の暮らしてきた地球は、古くから調査、研究がされてきた。近代科学以前から、ときには生活上の必要に迫られ、ときには知的好奇心によって、地面や山、川、海、風、地下を掘ると出てくる珍しい事物などが調べられてきた。

### 例：地球を測るデータ

地球そのものの大きさや形状・位置を測る測地学は生活上も不可欠で、そのはじまりは紀元前3世紀までさかのぼるともいわれる。現代では電波・レーザー光や人工衛星を使ったミリメートル単位の位置の決定が可能となった。交通や航法においてこうした精密測位技術活用が期待されている。

GPSなど衛星測位を使う座標系は地球全体で矛盾なく使えないといけないので、現代の位置情報サービスを活用するためには「世界測地系」と呼ばれる、グローバルな座標決定・利用のための基準システムが不可欠である。こうしたシステムを構築・

維持するために地球のデータを収集することは、少数の研究者ではままならない。

## 世界全体の座標基盤

こういう世界測地系といった考えの枠組みはすでに実現している。現在いくつかの世界測地系がある中でITRF (International Terrestrial Reference Frame) はほぼ世界標準だそうである。2002年、国土地理院などが主導し、我が国の測地系は、明治時代の測量に端を発する旧日本測地系からこの世界測地系に移行した。両者の距離差は日本付近だと400メートルほどとなる。

なお日常的に普及している測位サービスのGPSは、米国が維持する別の測地系で稼働しており、それ自体が日本測地系から400メートルほどずれているため、土地の基準点や印刷地図の緯度経度変更などを除けば、多くの地理情報サービスでは実用上はほぼ問題なかったらしい（つまり、すでにGPSの測地系自体がITRFと非常に近くなっていたためGPSサービスで見る日常の測位にはあまり影響がなかった）。

## 世界基盤を支えるボランティアなデータ事業

この世界測地系ITRFは、国際的に地球の測定にかかわる国際科学事業、国際科学データ事業であるIGS (International GNSS Service) やIERS (International Earth Rotation and Reference Systems Service), ILRS (International Laser Ranging Service), IVS (International VLBI Service for Geodesy and Astronomy), およびこれらを統括するともいってよいGGOS (Global Geodetic Observing System) などいくつかの活動によって支えられている（これらの機関は筆者が従事する国際科学データ事業World Data Systemに加盟して、より広範な国際的データ問題の議論に関与している）。各国の地理・測地機関や科学者グループによる測地データを収集・集積して、地球の重力場やプレート運動・地震・火山・海洋・気象あるいは地球深部構造の変

動などの影響で変化し続ける全世界の測地観測点の座標を計算して提供し続けている。

たとえば IGS は 100 カ国以上の大学・研究、現業機関等 200 カ所以上が運用する観測点の GNSS (GPS をはじめとする人工衛星による測位システム) データから国際基盤となるデータを提供している。各国で測地を行うのは各機関・科学者グループのボランティアでベストエフォートの貢献である。

一方、グローバルな活動が盛んな現代社会で、地球全体の整合的測地座標系は、市民、産業、政治、経済等に不可欠であるため、国連は 2015 年にグローバルな測地基準座標系の維持・推進を決議した。測地基準座標系維持や関連するデータ整備という公共事業(市場経済に従わないが社会に必要な事業)を行うことを、各国が国連加盟国として行っていく責務を負うことが明確になった。世界の科学者・研究機関ネットワークとそこから生まれるデータが支える現代社会のインフラの好例ではないかと思う。

## 地球科学データの今昔

### 16～17 世紀の地磁気データ共有

こうした高度なグローバルレベルの測地技術が存在しない時代には、たとえば、磁気コンパスによる方位測定は、海上の長距離航行における不可欠な手段だった。

図-2 は、1590～1699 年に測定収集された地磁気データの測定点約 12,000 点を示している。これ

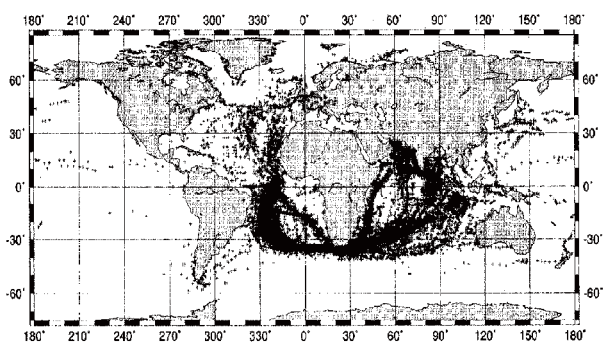


図-2 1590～1699 年に測定された約 12,000 点の地磁気データ (Jonkers et al., 2003)。

は欧州とインド、アジアの間を中心とした多数の航海で測定されたものである。

これらのデータ点は商業的・軍事的な航海で得たものもあるが、19 世紀中には何度かまとめられてデータテーブルの形にして書籍または論文におさめて出版されている。1 点 1 点のデータを保持していても価値はある程度以上に増加せず、社会的価値を増大させるには、データを集めてほかのデータの保持者と共有する必要があった。

つまり、ある一定の分野・業界内の必要性のために、あるいは業界内および周辺分野で活用されるために、紙媒体の形で数値データを出版する、つまり先駆的な形のデータ出版(データパブリケーション)が推進されていた、ともいえる。

### 研究分野の要請による国際データ共有

こうした先駆的なデータ共有、データ出版は、ほかの分野もまきこんで、徐々に国際枠組みとなっていく。たとえば、北極・南極を対象とする極地地方の国際共同観測計画「第 1 回国際極年」が 1882、1883 年に 11 カ国の共同事業として実施され、これは 1932～1933 年の第 2 回事業(44 カ国参加)へ引き継がれた。そして 1957～1958 年には国際地球観測年(IGY, International Geophysical Year)として、66 カ国が連携して世界中の 6,000 点の観測が行われた。得られたデータの国際管理体制の必要性が決議されて、世界データセンター体制(WDC, World Data Centre)が、日本学術会議も加盟する世界最大の国際アカデミー ICSU(国際科学会議)によって設立された。世界初の人工衛星、スプートニク打ち上げ成功(1957 年)とあいまって歴史的な宇宙時代の幕開けであった。

データが印刷物やマイクロフィルム等の形で保存・管理された時代から世界的なデータ整備・管理事業がはじまり、それは 60 年近く継続された。現在は技術革新によりインターネットと情報社会の時代となり、全学術分野を対象とする ICSU-World Data Sys-

tem (WDS) 事業とその国際委員会の設置が2008年に決議されて、WDC 事業は発展的に解消した。

### 先駆的なデータ解析システム構想

1970年代には、デジタルデータとその処理系のパワーを見て、研究者は新しいデータ利用の可能性を探り始めている。地球惑星科学の一分野である超高層物理学の国内コミュニティでは、WDCを含む関係研究組織を通信回線で結んでデータをやりとりして解析を行う総合解析システム「SOLTERTRON」が調査研究され、1975年の科研費報告書において提案された(図-3)。このように研究分野が自らの研究費を獲得してデータ解析のための統合的計算機システムを構想したことは、当時の通信回線ほか利用可能な技術や予算事情等を考えると大変先

進的かつ野心的(荒木徹, 2015)な試みであったといえる。

パイロットシステム開発には名大・東大・京大などが参加したが、予算事情そのほかにより残念ながらパイロットで終了せざるを得なかったそうである。ただこの思想は1990年代に「STEPNET」と呼ばれるプロジェクトへ引き継がれ、TCP/IP接続したUNIX計算機を、構築途上の学術情報センター(NACSIS)回線で接続して、接続された研究機関同士でデータを共有し、相互にそのデータを解析するシステムが開発されたとのことである。

### データ共有の今昔

幅広くデータ利用をしていく上で、データ公開、共有の考え方も変化してきた。本章ではデータ共有の在り方を見てみる。

#### 「従来型」データ共有

注意しなければならないのは、分野内のデータ共有では、その品質管理や配布、利用形態が、その分野内で最適化される傾向がどうしても見られる点である。昨今はFAIRデータ原則(Findable, Accessible, Interoperable, and Reusable)が国際的なデータ共有原則になりつつあるが、それ以前からも、公開・共有されたデータの利用、つまりデータの再利用(reuse)は行われていた。

データの再利用のためには、意識しなくてもユーザにとってFAIRの要素が満たされていた例は多いだろう。永続的識別子がなくても、学会内の研究者が、あのデータは〇〇研究所の誰々に聞けば分かる、などという共通理解があればそのときには発見可能である。解析のためのデータ特性の説明も同様で、測定原理やあり得る誤差要因、その性質などが学会内で共有されていれば(あるいはその論文情報が共有されていれば)、昨今のオープンサイエンスの議論に挙がるメタデータ(特にデータの特性に

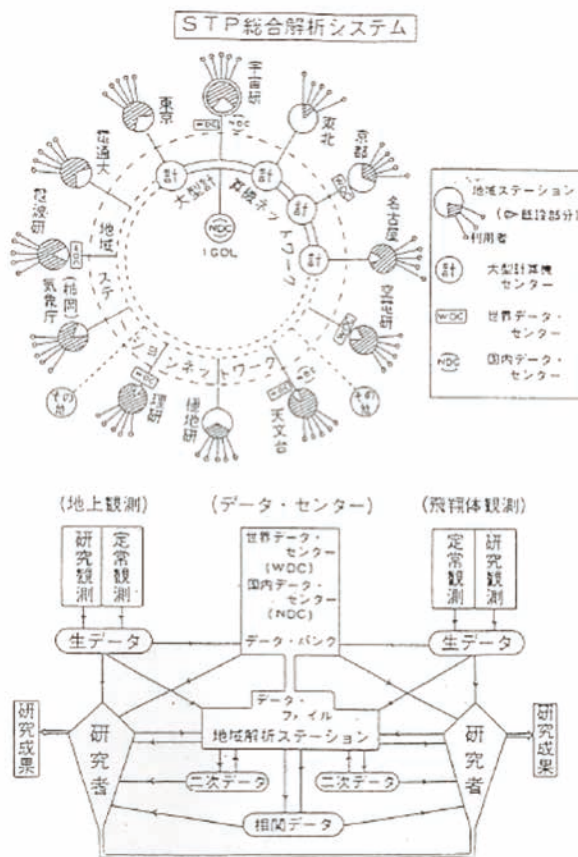


図-3 地球惑星科学の一分野(超高層物理学)における先駆的なデータ解析ネットワークシステムの構想。1975年の科研費報告書より(荒木 徹, WDS 国内シンポジウム, 2015)

する詳しいもの) や、データジャーナルに掲載されるデータ取り扱いに関する論説 (Springer Nature 社 “Scientific Data” 誌の掲載論文 Data Descriptor などのいわゆる「データ論文」) がなくても、既存の科学論文メディアとコミュニティ内の人伝えの流通で補えていれば、一定のデータ共有は分野内で成立していたといえるだろう。

### 新しいデータ共有の考え方

一方、近年のオープンサイエンスの文脈で議論されている「データ共有」では何が違うのだろうか。たとえば、欧州委員会が、欧州オープンサイエンスクラウド事業を立ち上げたときの議論を例にとると、分野横断研究の過去の失敗例、また欧州圏の経済を発展させるためのデジタル単一市場戦略 (Digital Single Market Strategy) において、データ自由流通、そのためのデータ基盤、また必要な標準化やツールの課題、等が重視された。

筆者の理解するところでは、たとえばこれまでの分野横断的研究が、異分野の研究者同志が知り合っ て顔が見える関係を構築して共同研究を行うような活動とすれば、他方では、いつでもアクセス可能なオープンアクセス論文誌の論文 PDF のように、研究者・専門家から好きなときに分野にかかわらず データを探してダウンロードしたり解析処理をして研究をすすめる、といった使い方で新しい成果を出せる環境、といったものが考えられる。

新しいデータ共有はそうした、個別の研究分野規範や文化、属人的情報流通に基づかなくても使える、共通の学術情報資産を形成することと言い換えられる要素が多分にある。誰でもいつでも、顔の見えない研究者同志で利用できるデータ資産にする、という考え方である。

データセットへの DOI 付与、データ生成者・整備者に連絡をとらなくてもデータを使えるようなメタデータの整備・ライセンスの整備、などなどが要請されてきているのだと理解できる。

## 印刷メディアから電子メディアへ

### データを取り扱う技術の変化

国際的なデータ共有が学会・研究者主導で行われてきた地球科学データの分野では、WDC 成立の当時、1950 年代では現場のデータアーカイブ (理学分野では「データセンター」ということが多い。今でいう「データリポジトリ」等) は、図書室のような本棚にデータブックやマイクロフィルムを収蔵するものが主だった。

その後計算機技術の発達はめざましく、20 世紀後半から 21 世紀初頭にかけて、特に社会の中のさまざまなシーンで、情報をデジタルで扱うことが飛躍的に増えた。

### 印刷メディアでの情報の取り扱い

図-4 は印刷テクノロジーとデジタルテクノロジーの簡単な年表を書いたものである。今の社会が享受しているイノベーション、テクノロジーは、たとえば鉄道、航空機、インターネット技術も、原則的に印刷媒体による科学・技術的知識の蓄積とその利用によって実現されてきた。

注意すべきことは、出版者や配送事業、小売り事業などの産業形態と経済効果の創出、また保存機関としての図書館や文書館の公的・私的設立や日本なら国会図書館法・著作権法など法制度をふくめた制度枠組みの構築、などが伴ってはじめて社会で真の

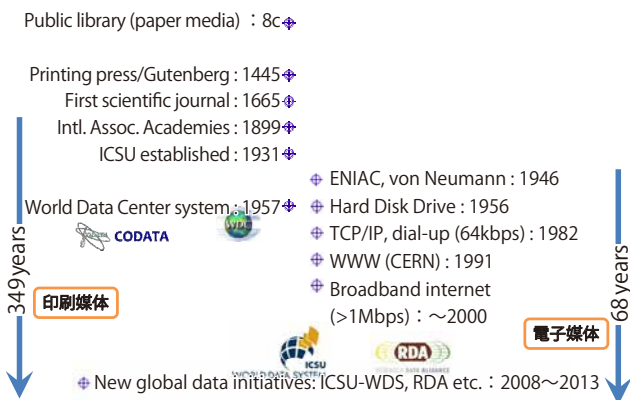


図-4 科学技術イノベーションの基盤としての情報共有の時代の流れ

インフラとして機能してきた側面があるということである。いわば、制度や文化面の利用・応用イノベーションあるいはソーシャル・テクノロジーといってもよい新たな基盤が構築され、印刷技術という工学的テクノロジーと結合系をつくることに成功した結果、数百年にわたって人類の知の基盤を構成してきたことである。

世界初の成功した学術ジャーナルは1665年にイギリスとフランスで発刊されたとされるが、英国王立協会のジャーナル第1号はいまもPDFファイルで読むことができる。インターネット普及までの300年以上、印刷媒体で保存されていたからでもある。

また読者の方は論文執筆時に海外の図書館にしかない論文が必要になり、複写を依頼した経験がおりかもしれない。これが想像以上に簡単な手続きで、筆者は約30年前の学生時代に簡単に photocopy が手元に届いて感激した記憶がある。大学図書館は、大学という機関の壁や国の壁をこえて、ちゃんと必要な情報資源をユーザへ届ける技術的仕組みと制度的仕組みをセットにしてサービスしてきた。筆者は、海外での同種の議論をまねて、これを「技術・社会結合系基盤」などと呼んでいる。

## 印刷メディアから電子メディアへの移行

一方、デジタルテクノロジーと電子媒体の歴史は100年未満であることは間違いないと思う（その歴史の正確な定義や評価は情報技術史の専門の方にご判断いただきたいが）。印刷メディアが三百数十年間維持発展させてきた社会的に必要な情報のアクセスと利用の基盤を実現できるかどうか、デジタルテクノロジーに問われる時代になりつつあると考えている。工学的・工業的テクノロジーと社会制度・文化・組織枠組みなどの結合系を構成し、経済活動のタイムスケールを超える何十年以上の（往々にして公的には明示されない）社会要請にこたえ続けることこそが、社会基盤として市民、産業、公的機能の要請にこたえて人類を支える技術と方法論になる

ための道であることを、印刷メディアの歴史が示しているのではないだろうか。

## デジタル社会基盤の未来へむけて

### 情報を共有・共用する基盤とは

図-5は、マイクロソフトリサーチ社と、英国バース大学が共同でRDA (Research Data Alliance) 総会で発表していたポスターからの図である (RDA 第2回総会, 2013年9月16日)。当時の同社副社長 Tony Hay 氏も参加して議論されていたが、多様な研究分野のコミュニティが図の中心にある「データ集約型研究」を実現していくためには、丸い図形の外縁部にある3つのキーワード、人的な面 (スキル、共同体制、オープン化の理解など)、環境的な面 (研究文化、法的・倫理的、商業的な要因など)、など多くの乗り越えるべき課題がある。残る1つの技術的な面については計算機・ソフトウェア基盤やそれを用いる共通実践事例の創出、またこれを活用するオープンネス、などが課題としてあがっている。

このポスターの議論は非常に幅広いコンセプトを扱っており、その目標への道程はいま議論されているオープンサイエンスのそれとほぼ同じといってよいと筆者は考えている。筆者が加筆した図中の日本語の吹き出しは、オープンサイエンス、あるいはデー

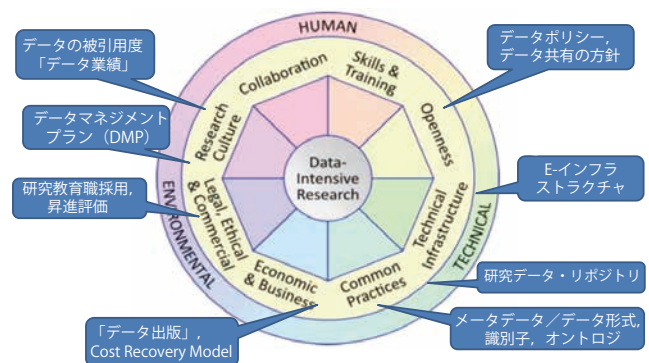


図-5 Research Data Alliance (RDA) Community Capability Model Interest Group (マイクロソフトリサーチ・英バース大学が共同主査) の概念説明図。図中の吹き出しは筆者による加筆。  
[https://www.rd-alliance.org/filedepot\\_download/383/230](https://www.rd-alliance.org/filedepot_download/383/230)

タ集約型科学研究を実現していくために今課題としてあがっている具体的な課題例、たとえばデータ引用・出版、データマネジメントプラン、データの研究業績評価への反映、ポリシー、データ基盤、リポジトリ、メタデータ・永続的識別子を示している。

## 「インターネット・オブ・データ」

もう1つ、海外の議論例を紹介したい。図-6はグローバル・デジタル・オブジェクト・クラウド(GDOC)と名付けられたネットワーク上のデータオブジェクト利用形態の構想を表している。インターネットの父といわれるRobert Kahn氏が提唱したデジタル・オブジェクト・アーキテクチャの考え方(Kahn and Wilensky, 1995)に基づく議論といわれる。IPアドレスやURLでなく、データ資源を物理的所在や所属集団などに依存しない識別子でマッピングしてサービスを実現する仮想空間を構想している。

ここでデータオブジェクトへのポイントとして想定されているのは永続的識別子「DOI」である。すなわちデータは、いつ消えてもおかしくないURL指定のWeb資源から、メタデータを伴った永続的識別子

が指定するデータ資源となることを意味する。原則的にはネット空間から消滅しないデータ資源と、物理的空間やインターネットの物理的接続などに依存しない、より抽象化されたサイバー空間中の情報資源として提供することを目指しているものと思われる。

この議論はResearch Data AllianceでRobert Kahn氏が会長を務めるCNRI(Corporation for National Research Initiatives)の副会長でもあるLarry Lannom氏から提出され、部会の1つData Fabric Interest Groupにおいても議論されている。

1970年代に計算機同士をつないだ特定分野の科学データ解析システムが構想されてから約50年を経て、さらに抽象化されたサイバー空間中で資源化された研究データの国際的基盤システムが構想されるにいたったことは感慨深い。今後50年後には私達のデジタル基盤は社会の中でどのようになっているだろうか。

(2019年2月25日受付)

村山泰啓 murayama@nict.go.jp

情報通信研究機構研究統括, ICSU-WDS 国際科学委員会 ex officio, 日本地球惑星科学連合理事, G7 オープンサイエンス WG 共同議長等。

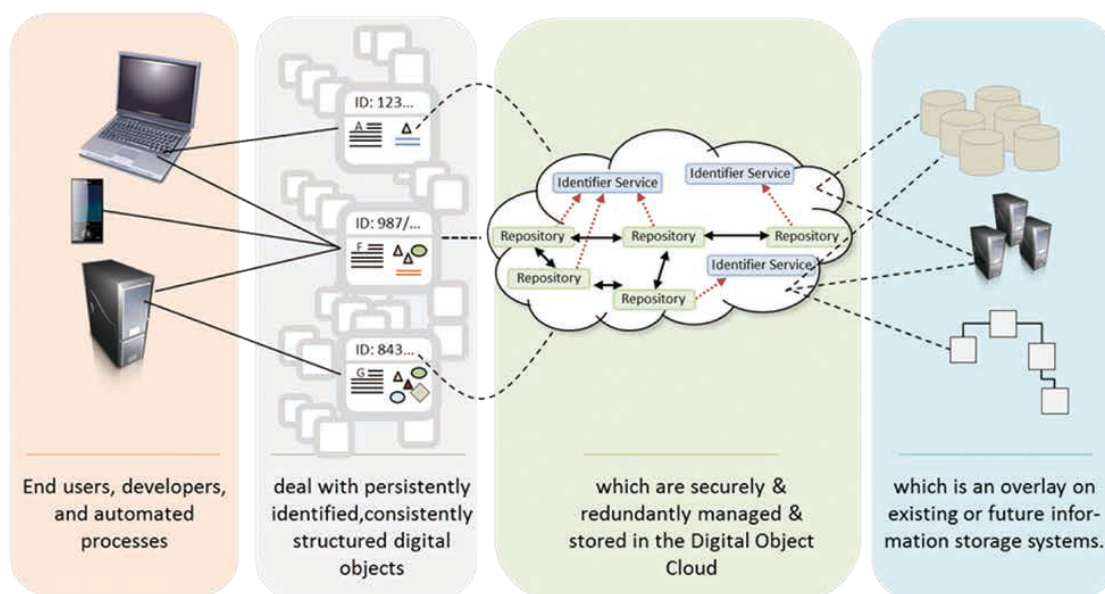


図-6 Robert Kahn 氏らが提唱するデジタル・オブジェクト・アーキテクチャに基づくグローバル・デジタル・オブジェクト・クラウド構想 (<https://www.rd-alliance.org/group/data-fabric-ig/wiki/global-digital-object-cloud>)