

予防医療を目指した医療情報統合

Medical Information Integration for Preventive Medicine

星野 隆之

要約 予防医療の実現のためには、情報が大きな役割を果たすことが期待されている。医療情報の活用においては、情報の標準化、検索の網羅性の確保、解析の容易さが必要である。また、体質や生活習慣に関する情報、およびバイオインフォマティクスや時空間情報など、医療情報を取り巻く様々な情報を統合して、推論し、仮説を検証するためには、体系的な知識表現が必要となる。このような条件を満たすため、プライバシー保護データマイニングとビッグデータを基本とするフレームワークを開発した。取り組み事例では、このフレームワークを適用することで、疾病予防などを目的とした医療情報の二次利用を迅速に進められることが確認できた。

Abstract Medical information is expected to play major role in realizing preventive medical care. To utilize medical information, standardization of information, securing the search coverage and ease of analysis are required. Also, systematic knowledge representation method is necessary to infer and to verify the hypothesis by integrating information such as body-build and lifestyle habit information, bioinformatics and spatio-temporal information that surrounding the medical information. To satisfy the requirements as stated, I have developed a framework which is based on the privacy protection data mining and big data. By applying this framework to our case example, we quickly proceeded with the secondary use of medical care information aimed at disease prevention.

1. はじめに

日本は、世界最高水準の平均寿命を達成している。しかし、高齢者の慢性疾患による受療、疾病の罹患率、要介護率が増しており、「平均寿命と健康寿命との差の短縮」が望まれている^[1]。病気を超早期に発見し、発症前の予防法を確立することは、平均寿命と健康寿命の差を短縮させること、つまり「健康長寿社会を構築する」ために不可欠であるといえる^[2]。

日本ユニシスは予防医療を支援・進化させるためのICT利用について、2010年度より取り組んできた。本稿では、2章で予防医療における医療情報統合の意義と課題について整理し、3章で医療情報を統合管理するためのフレームワークを提示する。また4章で、2010年度以降の具体的な取り組みと成果を紹介し、現在までに実証してきた技術の状況とこれからの展望を述べる。

2. 予防医療における情報統合の意義

予防医療では、病気の予兆を検知し、病気が重篤化する前に適切な処置を行うことを目指している。大部分の病気は「遺伝的な要因」と「環境的な要因」によって発症するため、両者を正確に把握・分析し、「遺伝的要因と発症し易い病気の種類、その発症に関連する環境的要因」

を解明することが必要となる。

環境的要因のうち生活習慣を把握する手段としては、デジタルな情報を利用する取り組みが2010年度から進められている^[3]。また、遺伝的要因を示す体質については、客観的に表現し共有することはできないとされてきたが、ヒトゲノム解析の進展によって、個人が持つ体質をゲノム差異によって表現できるという仮説が導入され、個人の遺伝情報に応じた医療の基盤が作られつつある。また、分子レベルでのバイオマーカーを利用することで、臨床症状が現れる前に先制的な治療等を施せる可能性も生まれている。

2.1 予防医療における情報の役割

予防医療の実現に際しては、次の点から情報が大きな役割を果たすことが期待されている。

第一は、ゲノム情報がデジタル性を持つ点である。ゲノムは遺伝物質の総称で、24種類の染色体に分散して情報が蓄えられている。この情報は科学物質であるDNA（デオキシリボ核酸）が担い、4種類の塩基の配列として表現できる。

第二は、ゲノム差異と疾病の確率的関係の把握に、情報が大きな役割を果たすという点である。ゲノム差異が生物としての個体に及ぼす影響は確率的なものであり、一般的に疾患は、最終的に同じ症状を示す病気であっても、体質と生活習慣やその他の環境要因が複雑に絡み合っただ様な経過をたどる。分子疫学コホート研究では、研究上重要なこの情報を、記憶に頼るのではなく、一定の研究参加者集団を長期的に追跡研究することで、正確に把握し、精度の高い研究にできるとしている。そのためには、膨大な量のゲノム、バイオマーカー、生活習慣、病歴情報を収集し、構築された統合データベースを利用して研究を進める必要が生じてくる。予防医療のためには、このような高精度かつ大規模にデータを収集するための、分子疫学コホートが必要である。

第三に、この分子疫学コホートには臨床記録などの医療情報が必要なことである。疾病サンプルとコントロールサンプルの収集にあたっては、さまざまな検査の結果や家族歴、合併症の有無についての詳細な情報が必要になる。しかし、現実の医療の現場では、必ずしも詳細な医療情報が得られているわけではない。このため、分子疫学コホートを長期的に運用し、継続的なデータ収集を行い、詳細な医療情報の統合を目的とした、従来とは異なる特性を持つ医療情報二次データベースを構築することが求められる。

2.2 医療情報の二次利用と分子疫学コホート

本節では、医療情報の二次利用、関連する情報の統合、データの集積という点から、その特性と課題を紹介する。

医療情報の二次利用は、従来より、病院の経営改善や医療の質の向上という観点から必要性が問いかけていた。医療の質向上を目的とした医療情報の二次利用形態には、医療現場での個々の診療や検査などを記録したEMR（Electronic Medical Record）を一次データとして、診療機関をまたがって患者ごとに病態を管理できるEHR（Electronic Health Record）や、一人一人が自分自身の診療や健診の記録を管理・閲覧できるPHR（Personal Health Record）などがある。

この他にも、医療情報は、医学領域における先進的な基礎研究成果を産業化まで進める過程において様々な形で利用される。

分子疫学では、あらゆる医療情報や生活情報を、膨大なゲノムの配列情報としての「遺伝子型」に対する「表現型」として捉え、ある集団のゲノム中に見られる変異と、その集団の医療情報・家族歴・生活情報などを関連付け、両者の対応関係の有無や程度を算出することが主要な研究方法になる。このためには、情報の標準化、検索の網羅性確保、解析の容易さが求められる。

しかし、従来の情報システムは、個々の医療機関で蓄積されたEMRからの情報の抽出が主であり、二次利用に必要な複数の機関の情報システムからの情報の取り込みや統合が難しい。また、研究における医療情報の二次利用独特の倫理的な困難もある。通常、医療情報を提供するときには、本人が研究内容やリスクについて承諾する必要があるが、研究という行為の性質上、いかなる研究に用いられるかを、その時点では全て提示することは難しい。そのため、通常の疫学コホートに求められる以上に、研究の進展を支える情報の扱いの自由を確保することと、参加者の同意方式についての配慮が必要となる。

3. 予防医療に関連する情報の統合と集積

対象とする情報システムを取り囲む環境を視野に入れたデータの構造を、データアーキテクチャ(図1)に示し、詳細を本章の各節で説明する。

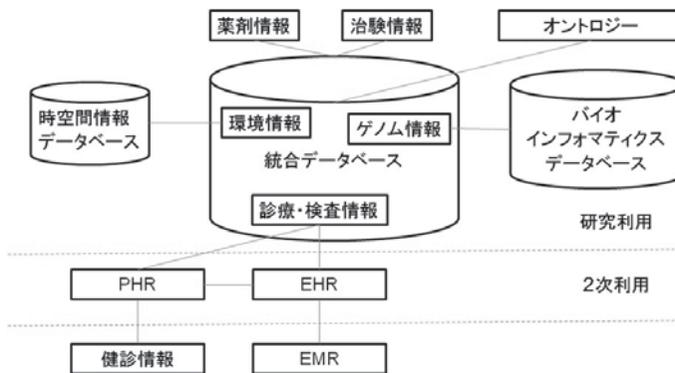


図1 データアーキテクチャ^[4]

3.1 情報の統合

分子疫学コホートの考えに基づいて、医療情報システムで扱うデータをモデル化すると、化学物質間の関係と、臨床記録から得られる生理的、生物学的反応との関係に要約できる。化学物質間の関係では、人間に薬を与える処方、人間を構成する遺伝子、血液、組織などを調べる臨床検査の結果と、放射線など環境的な物質、人間が摂取する食べ物などの物質が対象となる。そして、物質に関わる情報を保存、集積、加工することで、意味を導出する視点が基本となる。

このような視点に基づく研究にはバイオインフォマティクスに関するデータベースが必要となる。例えば、京都大学化学研究所が維持しているKEGGでは、遺伝子を軸にタンパク質が相互作用する薬剤や化学物質、化学反応や疾病なども網羅した統合的なデータベースになっている。しかし、生命科学やバイオインフォマティクスの発展は速く、新しい実験法と新しい種類のデータが生み出され、データベースの統合は完了していない。

環境情報を分析するには、時空間情報のデータベースが必要となる。時空間データについては、ISO/TC211などの国際標準化が進んでいるが、日本の国土全体の時空間データを把握する基盤は整備途上である。

これらの多様で発展を続ける情報を統合して、並び合わせ、推論したり、仮説を検証したりするためには、疾患や解剖学的な構造などに関する体系的な知識表現（オントロジー）が望まれる。これについては、SNOMED-CTによる診断情報の記述など、医療情報を記述する際に用いる共通語彙（用語とコード）が整備され、形式化や標準化がすすめられている。

3.2 情報の集積と利用

現在の医療情報システムでは、二次的な情報は、行政圏、医療圏、経済圏などに依存して収集されているが、日本の医療は、「いつでも、どこでも」病院にかかれるフリーアクセス制であるため、一元的に情報を捉えることが難しい。分子疫学コホートの対象とするデータの寿命が数十年であることを考えると、生涯健康管理や臨床研究への応用ができるよう、長期にわたるデータの継続性が確保でき、診療支援、臨床支援、経営分析に活用できるような新たな情報システムが不可欠となる。

具体的には、追跡調査のために、疾患発生や動態情報の把握方法の確立、基幹病院との連携体制の確立、オントロジーの統一、共通するIDを前提としたシステムの設計が必要となる。そのためには、医療データの二次利用を主眼に置いた汎用的なデータモデルを提供することで、広範囲なデータを蓄積できるようにし、基幹となる情報システムのベンダに依存しないデータ抽出を可能にすることが求められる。

また、大規模データを継続的に管理するためには、データの拡張性、可用性が必要となり、さらに、多くの医療機関のデータを低コストで集積・分析できる必要がある。ここでは情報技術の発展を考慮に入れることが望まれる。現在の情報技術では、インターネットに見られるように、個々のコンピュータをノードとしたリンク構造でITネットワークを構成し、処理するノードの数を増やすことで、全体の性能を向上させていくスケールアウトというアプローチが主流である。このような医療情報の二次活用に必要な特性を満たすためには、どのようにデータを分散させるかが重要であり、その点で基本となる技術は、ビッグデータとプライバシー保護データマイニングである。

3.3 医療情報の二次利用における情報の構造と流れ

医療情報の二次利用における情報の構造と流れは、電子商取引と同様である。日本ユニシスでは、このような情報フローを実現するには、メッセージの構造と交換の方式、分散した情報源へのアクセスと制御を定めたフレームワークが必要と考えている^[5]。

一方、複数の医療機関にまたがった患者診療/健診などの記録を蓄積し続けることによるデータの増加に伴い、データのロードおよび検索時間が増大するおそれがある。近年、クラウドコンピューティングを活用した企業群（グーグル、フェイスブック、ヤフーなど）から発達してきたビッグデータ技術は、以下の点から、このような特性に適合している。

① メッセージ構造：

記述の自由さと検索の容易さを両立させた半構造化データ。

② 交換の方式：

さまざまな情報源から集積されたデータが、標準化されたデータとして横断的、時系列的に利用できる。

③ 分散した情報源：

画像や音声は、個々の記録がオブジェクトとなった階層として表現される。

さらに、これらを支えるのが分散ファイルシステムである。分散ファイルシステムは、複数のノードのディスクを組み合わせて一つのファイルシステムとして見せる技術の一つであり、スケールアウトによって拡張でき、ノード数を増やしても局所的な入出力性能を維持できる。

電子商取引のフレームワークをこのようなビッグデータの技術で拡張し、医療情報の取り扱いを考慮したフレームワークを図2に示す。

このフレームワークに基づいた実際の事例については、次章で紹介する。

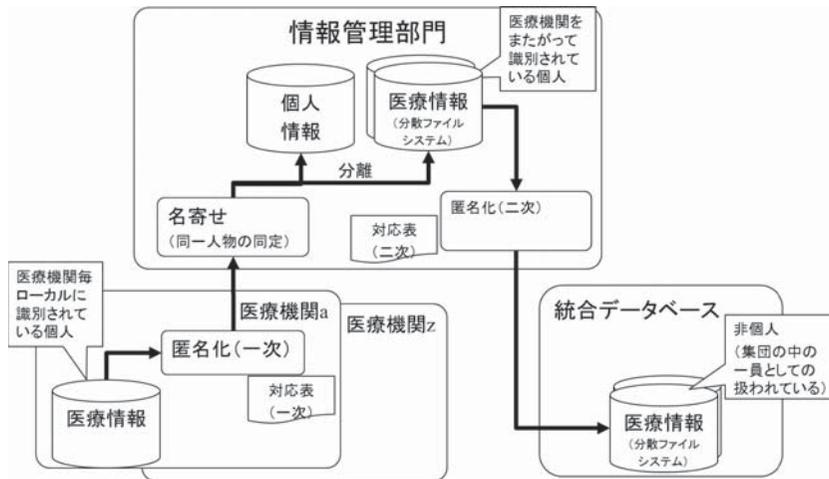


図2 医療情報統合のためのフレームワーク^[4]

4. 取り組み事例

4.1 生涯カルテを実現する医療情報統合基盤

地域医療連携への取り組みとして、佐渡地域医療圏の地域医療再生計画の一環として構築された「佐渡地域医療連携ネットワークシステム（愛称「さどひまわりネット」）がある^[6]。日本ユニシスが2012年3月に開発パートナーとして当該システムの開発を受注し、2013年4月に第一期本番、同12月に全面本番を迎えたもので、生涯カルテを実現する医療情報の統合基盤として位置づけられる（図3）。

データ源泉である医療機関等でのデータ整備の負担を極小化し、地域における医療情報の共有を実現している。

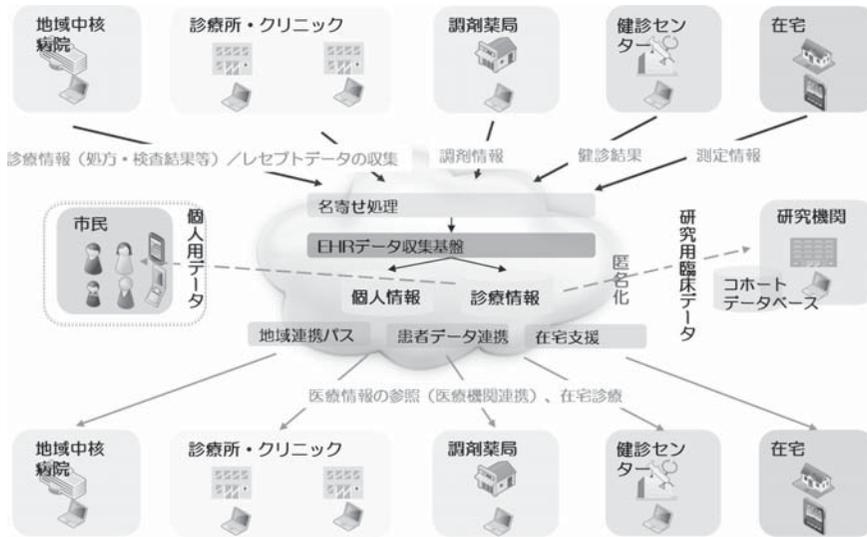


図3 医療情報統合基盤 概要図

4.2 地域 EHR システム

地域 EHR システムは、複数の医療機関の診療データや検診機関の健診データを長期（数十年）にわたり蓄積・共有し、地域の医療機関での医療情報連携、および生活習慣病のコホート研究に利用するための基盤システムである。複数の地域において、病院・コホート研究機関を主体に、自治体・NPO と連携し、日本ユニシスが開発パートナーとして 2010 年にプロジェクトがスタートし、システムの構築を終え、運用を続けている。基本構造を図4に示す。

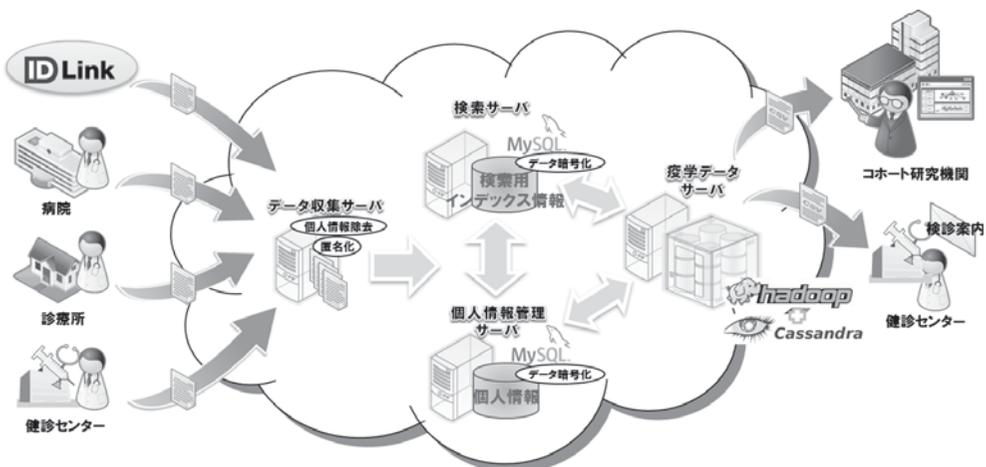


図4 地域 EHR システムの基本構造

この取り組みでは、非構造化・複雑・大量データの扱いと、データの二次利用・研究利用における個人情報の扱い、プライバシー保護に関する基本的な技術を実現した^{[7][8]}。

4.3 ゲノムコホート研究のための統合情報基盤

日本ユニシスは、京都大学医学研究科附属ゲノム医学センターと、ゲノムコホート研究のための情報基盤開発に2011年より取り組んでいる^{[8][9]}。同基盤は、予防医学にむけた疾患発症追跡のための研究基盤として開発を始め、2012年より運用を開始した。以降、全国レベルでの展開を視野に、高品質かつ安全にデータや試料の収集・統合・蓄積・提供を可能とする機能を整備している。

対象となるデータは、保険者である地方自治体もつ健診情報、電子カルテシステムや部門システムにあるEMR、地域連携パスなどを通じて得られる診療情報である。これらを、研究計画（プロトコル）で定義されたメタデータ（データの構造、属性に関する情報）に基づき収集・統合し、個人情報保護への対応を行い、研究プロジェクトへ提供している（図5）。

この統合情報基盤では、地域EHRシステムでの実績を基に、以下の機能を中心に、技術を洗練させ、実運用を行っている。

- ① 多段階匿名化と個人情報秘匿
- ② 医療機関、健診センターにまたがった患者（個人）の同定
- ③ 検索－病名、症状、検査結果による研究対象の特定

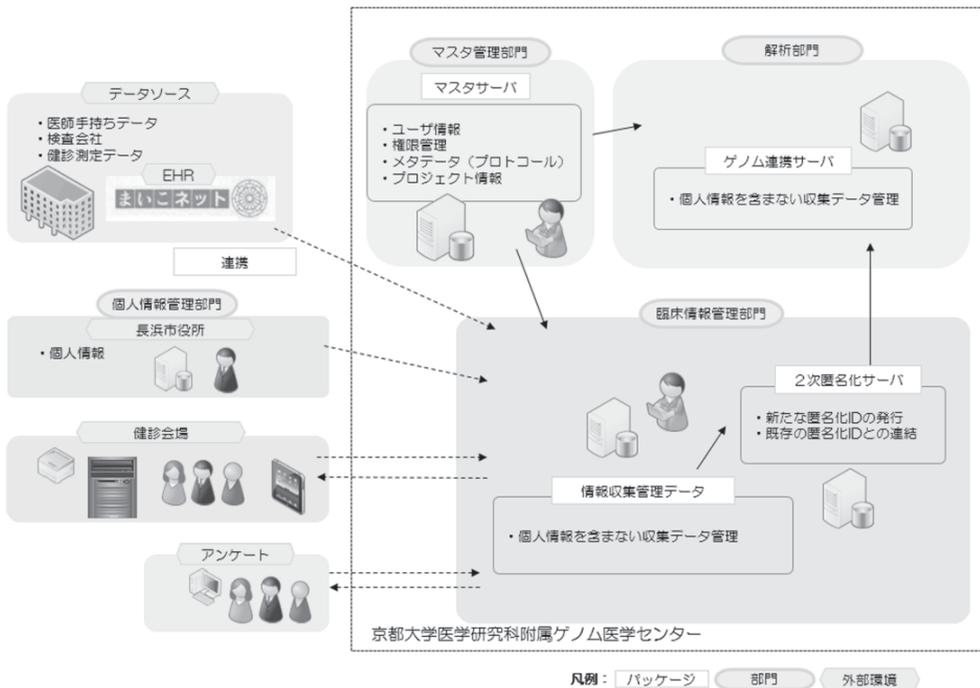


図5 ゲノムコホート研究のための統合情報基盤 概要図

5. おわりに

取り組み事例では、ビッグデータ技術の採用によるEHRデータの拡張性・可用性、標準化データを利用し、クラウドコンピューティングを活用した設計による低コストでのデータの集積と分析、疫学データに必要とされるデータウェアハウスのような時系列性を実現している。本フレームワークを他の疾患などに適用することを通じて、医療情報の二次利用や研究利用に

よる予防医療の進展に貢献していきたい。

また、京都大学が取り組みを予定している「低温冷凍システム (CAS*)」の研究開発/導入」では、CAS 冷凍技術を用いた生体試料 (細胞, 血液, 臓器など) の凍結・長期保存研究から新たなデータが得られることが期待でき、大量の生体試料そのものと関連する電子データとの統合化技術や、データ解析技術の更なる洗練が必要になると考えている。更に、「多重匿名化」や「プライバシー保護とトレーサビリティの両立」など新たに発生する課題について、あるべき姿の明確化と実現に関する検討を京大ゲノム医学センターと共同で開始することを想定している。

今回紹介した医療情報統合のフレームワーク、事例が、医療情報はもとより、ビッグデータの利活用のための情報統合・解析基盤の参考になれば幸いである。

最後に、今回紹介した情報統合・解析基盤の研究開発に助言・支援をくださった方々、および各事例に取り組み・情報を提供いただいた方々に、深くお礼を申し上げる。

* 1 CAS (Cells Alive System). 株式会社アビーが開発した細胞を生かす技術で、凍結装置と組み合わせることによって細胞破壊を最小限にとどめ、凍結時の鮮度・風味・旨味を維持する技術。

- 参考文献**
- [1] 「健康・医療戦略 改訂 2014」, 2014 年 7 月 22 日閣議決定
 - [2] 提言「100 万人ゲノムコホート研究の実施に向けて」, 日本学術会議 ゲノムコホート研究体制検討分科会, 2013 年 7 月
 - [3] 新堀聡, 「予防医療を実現する生体情報収集基盤」, ユニシス技報, 日本ユニシス, Vol.30 No.4 通巻 107 号, 2011 年 2 月
 - [4] 羽田昭裕, 「個別化医療を目指した統合情報基盤」, 細胞, ニューサイエンス社, 46 巻 14 号通巻 617 号, 2014 年 12 月
 - [5] 大沼保夫, 「企業における E ビジネスのための次世代情報交換システム」, ユニシス技報, 日本ユニシス, Vol.21 No.2 通巻 70 号, 2001 年 8 月, (特許 3632845, 特許 3588061)
 - [6] 渡辺和彦, 田中孝治, 「地域完結型医療を実現する情報集約型医療連携ネットワークの構築」, ユニシス技報, 日本ユニシス, Vol.34 No.2 通巻 121 号, 2014 年 9 月
 - [7] 星野隆之, 「大量かつ複雑な非構造化データを扱う解析基盤の仕組み」, ユニシス技報, 日本ユニシス, Vol.31 No.4 通巻 111 号, 2012 年 3 月
 - [8] 沖俊吾, 「医療システムでの非構造化データ活用事例」, ユニシス技報, 日本ユニシス, Vol.31 No.4 通巻 111 号, 2012 年 3 月
 - [9] 福田健太, 「ヒト生命情報統合研究を支える ICT 活用」, ユニシス技報, 日本ユニシス, Vol.34 No.2 通巻 121 号 2014 年 9 月
 - [10] 沖俊吾, 「大規模ゲノム疫学研究の統合情報基盤の構築事例」, ユニシス技報, 日本ユニシス, Vol.34 No.3 通巻 122 号, 2014 年 12 月

執筆者紹介 星野 隆之 (Takayuki Hoshino)

1985 年日本ユニシス(株)入社。人材育成部門, データ利用技術部門, 電力事業部門で情報系のシステム開発プロジェクトに従事。2006 年より R&D 部門において, データエンジニアリング関連の研究開発プロジェクトに従事。現在, 総合技術研究所イノベーション室に所属。PMAJ 公認 PM レジスタード。

