

合	同	分	科	会	選	出
---	---	---	---	---	---	---

合同分科会 2012 年度会合 より

ペタバイトのデータの海に
ヒッグス粒子を追う
- LHC アトラス実験の物理解析 -

坂本 宏
(東京大学)

ペタバイトのデータの海にヒッグス粒子を追う

LHC アトラス実験の物理解析

坂本 宏

東京大学 素粒子物理国際研究センター

[アブストラクト]

ジュネーブ郊外にあるCERN研究所で稼働中のLHC(Large Hadron Collider:大型ハドロン衝突型加速器)はヒッグス粒子と思われる新粒子の発見などの成果を上げている。我々はアトラス実験に参加しデータ解析を行っている。LHC実験のデータ解析は世界各国の100を超える大学・研究機関を結んだコンピューティンググリッド、通称WLCG上で行われている。日本では東京大学に設置された地域解析センターが数千のCPUコア、数ペタバイトのディスク容量などの計算資源を提供し解析に貢献している。

[キーワード]

衝突型加速器、素粒子物理学、ヒッグス粒子、グリッドコンピューティング、世界分散解析網

1. はじめに

2012 年 7 月 4 日、ジュネーブ郊外にある CERN(欧州合同原子核研究機構)[1]で記者発表が行われ、126GeV(ギガエレクトロンボルト)付近の質量を持つ新しい粒子が発見されたことが報告された。この粒子は、我々が長年探し続けてきたヒッグス粒子と考える。ヒッグス粒子は素粒子の標準模型と呼ばれる理論において唯一未発見の粒子であった。標準模型は、物質を構成する粒子であるクォーク(陽子や中性子の構成要素とその仲間)とレプトン(電子の仲間)、力の媒介をするボソン、そして粒子に質量を与えるヒッグス機構がもたらすヒッグス粒子を予測、素粒子現象を非常に高い精度で記述できる。一方、この理論はいくつもの重要な事項を説明していない点で不完全でもある。ヒッグス粒子の性質の詳細な調査から、標準模型を超える理論への手がかりが得られると期待されている。

素粒子の性質を調べるためにはそれらを作り出すだけ十分なエネルギーを生成するための加速器が必要である。量子力学が示すように粒子は波動性を持つ。その波長は粒子のエネルギーに対応する。光学顕微鏡では見えない(可視光の波長より小さい)ものを見るため、より波長の短い電子を用いた電子顕微鏡が作られたように、より小さい素粒子を観測するにはより高いエネルギーの加速器が必要である。エネルギーフロンティアと呼ばれているが、粒子の束を可能な限り高いエネルギーに加速し、相互に正面衝突させることにより非常に高いエネルギー状態を作り、そこで生成された粒子の性質を調べる。より高いエネルギー状態を作ると言うことは、宇宙の始まりであるビッグバン直後に作られたであろう高エネルギー状態に近づくという意味を持つ。現在到達しているエネルギーは、宇宙開闢後 1 ピコ秒(10^{-12} 乗秒)に相当する。エネルギーフロンティアは我々の宇宙がなぜこのような姿なのかを解き明かす道でもある。

一般に重い粒子はすぐに軽い粒子に崩壊すると考えられている。そのため、生成された重い粒子の性質を直接調べることは出来ず、その崩壊生成物から元の粒子をたどる。崩壊粒子を一つ残らず検出するため、衝突点の周囲をほぼ完全に検出器で覆い尽くす。崩壊粒子のエネルギーや運動量などを正確に測定するため、検出器には高い分解能が求められる。そのことはセンサーの寸法が小さくなる、センサーの数が膨大になるこ

とに相当する。同時に、粒子の高いエネルギーを正確に測定するためにはそれだけ分厚いセンサーの厚みも必要になる。このようにエネルギーフロンティアでは巨大な検出器が必要になる。

非常に発生確率の低い現象を調べるため、粒子束の衝突は極めて高い繰り返し周期で起こる。このことはセンサーからのデータを非常に短い時間に読み出すことを要求する。そのためには専用の電子回路系が必要であり、検出器は膨大な数の ASIC(特定用途向け大規模集積回路)と光ファイバーに覆われていると言ってよい。その結果、読み出されたデータを処理するのに必要となる計算資源は空前の規模になる。このようにエネルギーフロンティア実験は最先端技術の集約によって初めて実現すると言ってよい。この一つの例として LHC 加速器アトラス実験を取り上げて見ていこう。

2. LHC 加速器とアトラス実験 ~ データ解析への要請

LHC[2](Large Hadron Collider: 大型ハドロン衝突型加速器)はジュネーブ郊外、スイスとフランスの国境をまたぐ CERN に建設された陽子・陽子衝突型加速器である。地下約 100m に掘られたリング状トンネルは周長が約 27km ある。その中に 1,600 を超える超伝導電磁石が並べられ、高周波加速空洞により陽子を 7TeV(テラエレクトロンボルト、10 の 12 乗電子ボルト)まで加速する。約 1 千億個の陽子の塊(バンチ)を右回り、左回りそれぞれの加速管に入射加速し、リング上の 4 か所で衝突させる。バンチは 25 ナノ秒毎に交差する。アトラス[3]はその衝突点の一つに置かれた巨大な検出器である。

アトラス検出器は数多くの種類のセンサーと 2 種類の電磁石から構成されている。電磁石は放出された荷電粒子の運動量を測定するためのもので、中心部をソレノイド電磁石、さらにその外側をトロイド電磁石が覆っている。センサー部は大別して、中心部で荷電粒子の飛跡を検出する内部検出器、その外側で粒子のエネルギーを測定するカロリメータ、さらにその外側まで貫通するミュー粒子を測定するためのミューオンスペクトロメータからなる。内部検出器は高い位置分解能を持つ、薄いセンサーからなる。半導体検出器が主に使われている。カロリメータは、内側が鉛と液体アルゴンを媒体とした電離検出器で、電子やガンマ線が起こす電磁シャワーを検出する。その外側に燐光を発するシンチレータと鉄のサンドイッチからなり、原子核反応によるシャワーを検出する。これらにより、粒子の全エネルギーを測定する。ミューオンスペクトロメータはさらにその外側に配置され、カロリメータをすり抜けてくるミュー粒子を測定するガスを用いたセンサーからなる。アトラス検出器全体の大きさは高さ幅約 24m、長さ 44m に及び、重量は 7,500 トンと見積もられている。

アトラス検出器のセンサーから読み出される信号チャンネル数は 1 億近い。そこからビーム交差毎、つまり 25 ナノ秒毎にデータは読み出される。あまりに膨大なデータが読み出されるため、そのまますべてを記録することは出来ない。そこで 40MHz(周期 25 ナノ秒)でサンプルされたデータのうち、物理的に興味深いものを選んで記録することになる。このトリガーと呼ばれるプロセスは 3 段階で実現する。まずハードウェア論理回路により 40MHz のデータから興味深そうな 75kHz 分のデータを抽出する。続いて二段階に分けて PC ファームでデータを解析し、最終的には 400Hz 程度まで絞る。1 事象のデータサイズは圧縮後で約 1MB あり、400MB/s の速度でデータが送出される。1 年間に蓄積される生データ量はこの結果、約 1/3 稼働する(年間 1 千万秒)として 4PB(ペタバイト、10 の 15 乗バイト)に及び。

この生データはセンサーからのデジタルデータであり、そのままでは物理解析には適しない。センサーのチャンネル識別番号やデジタルデータパターンからなる生データを解読して物理的に意味のある情報に変換していく必要がある。まず、位置検出器では信号データを空間座標に変換する。多数の空間座標データに対してパターン認識を行い、荷電粒子の飛跡を再構成する。再構成された飛跡と、該当する空間の磁場情報から粒子の運動量などを決定する。カロリメータの場合、信号データを位置情報とその点でのエネルギー投下量に変換する。空間的に近接したエネルギー投下量を集約してエネルギーの塊(クラスター)を再構成する。再構成された飛跡やクラスターを総合して、その粒子がクォーク由来のもの(ハドロンと呼ばれる)かレプトンかなど種類の識別を行う。この段階で初めて、素粒子反応の事象情報として物理解析に使いやすい形

になる。取得された生データから粒子情報を取り出すプロセスを再構成プロセスと呼ぶ。

素粒子反応によって作られた重い粒子は複数の軽い粒子に崩壊する。得られた粒子情報から崩壊元の重い粒子を予測するのだが、これは一意的に決まるものではない。元の親粒子としていろいろな可能性が考えられる。素粒子反応は確率的な過程であり、それ故、親粒子の可能性を統計的に予測するしか方法がない。新粒子の存在を探索するためには、その粒子を含め、様々な粒子をモンテカルロシミュレーションによって生成し、それらがアトラス検出器でどのように検出されるか、計算機上で模擬実験を行う。その模擬実験のデータと実際に取得されたデータを比較することにより、新粒子の存在など様々な仮説の検証を行うことが出来る。この模擬実験は、検出器を構成する多数のセンサーや構造体などの中で粒子がどのように反応するかを一つ一つシミュレートしていく。そのために膨大な計算時間が必要となる。

データ解析への要請としては、次のようにまとめられる。

- ・ 実験後大きく遅れることなく、生データを再構成し物理解析データを作る
- ・ 実験データとほぼ同程度の事象数の模擬実験データを生成する
- ・ 模擬実験データも生データと同様に再構成し物理解析データを作る
- ・ 生、模擬実験とも各段階のデータファイルを安全に保管管理する
- ・ 各段階のファイルを効率的に読み出し、物理解析プログラムを走らせる

数量的には、生データが年間 4PB 程度生成されるが、模擬実験データも同様な量に及ぶ。また、再構成プロセスによって生成される物理解析用データや解析結果もファイルとして保管される。効率的に解析を行うためには同じデータの複数個のコピー(レプリカ)を用意する。それにより、万一の時のデータ損失も避けられる。その結果、取り扱われるべきデータサイズは 100PB に及び、ファイルの数も 1 億個に及ぶ。

再構成プロセスは、1 事象処理するために標準的な PC で 40 秒程度かかる。毎秒 400 事象収集されるデータを遅れず再構成するためには 16,000 個の CPU コアが必要になる。また、1 事象の模擬実験には 1,000 秒以上かかる。生データと同程度の事象を生成するには、160,000 個の CPU コアが必要になる。このようにつてない規模の計算資源が必要となる。これはもはや単一の研究機関で維持できる資源量ではない。

3. 世界 LHC 解析網 ~ 世界最初の実用配備コンピューティンググリッド

見てきたように、LHC での実験データ解析には膨大な計算資源が必要であり、それを実現するためには国際的な解析網を構築する必要がある。CERN では MONARC [4]と名付けられたプロジェクトが 1998 年に立ち上げられ、LHC 実験データ解析基盤の基本設計が与えられた。それによれば、高速国際ネットワーク接続を前提に、CERN を Tier(ティア:階層)0 とし、最初の再構成を行い、データを送出する。世界に 10 程度の Tier1 センターを設置し、生データの保存と再度の再構成を行う。各国に 1 つ以上の Tier2 センターを設置し、物理解析を行うとともに、模擬実験データ生成を行う。各研究機関には Tier3 施設が置かれ、ユーザ解析が行われる。このモデルに基づき、各実験グループはそれぞれのコンピューティングモデルを確定した。

次にそのモデルを実現するための技術として何が使えるか。100 を超える世界中の研究機関を国際ネットワークで接続し、データをよどみなく流し、効率よくプログラムを走らせる必要がある。そのためのインフラとして注目されたのが当時まだ実証試験段階であったコンピューティンググリッド[5]である。これは地理的に分散した多数の計算資源をネットワーク接続し、その上で実行されるミドルウェアにより統一的に資源管理し、あたかも単一の計算システムのように利用者に提供するものである。利用者はグリッドミドルウェアと対話することにより、データや CPU がどこにあるかを気にすることなく最大限資源を利用出来る。グリッドの用語は、自宅のコンセントから電気を使うとき、それがどこの発電所で発電され、どこを通過して供給されたか(パワーグリッド)など全く気にしないことと同じように計算資源を必要に応じて使える環境を目指して与えられた。

LHC の実験グループは共同でグリッドミドルウェアの開発と配備のためのプロジェクトを開始した。これが

WLCG(Worldwide LHC Computing Grid)[6]である。WLCG プロジェクトは 2001 年に開始された。ミドルウェア開発は実際には単一のフレーバーを用いるのではなく、参加国の状況などを反映し、欧州を中心とした EGEE[7]の gLite[8]、北欧中心の Nordugrid[9]の ARC[10]、米国中心の OSG[11]というように複数のミドルウェアが相互乗り入れをしながら採用された。EGEE は、現在は EGI[12]に、gLite は EMI[13]に引き継がれている。

グリッドミドルウェアは主に次のような機能を実装している。まずデータに関して、レプリカカタログなどデータ管理を行う。また、効率的にファイルを転送するデータ転送サービス機能を持つ。これらを実装するために、サイトの記録資源をアクセスするためのインターフェースが必要である。テープ装置をバックエンドに持つ階層化記録も採用されている。計算資源に関しては、ローカルバッチシステムとのインターフェースが用意されている。また、適切な計算資源の発見と負荷分散を担うワークロード管理機能を持つ。ジョブ実行状況を管理・記録するログ・ブックキーピング機能も用意されている。サイトの資源量や構成などの情報をやりとりする情報管理システムが基盤にあり、サイトやデータ転送など様々な情報に関するモニター機能が用意されている。ユーザやホストの認証には X.509 認証機構が用いられ、仮想組織メンバーシップによりグリッド資源へのアクセスが制御されている。

グリッドミドルウェアの大規模な試験配備は 2004 年頃から始まり、様々なスケールテストを通してミドルウェアの改良が続けられてきた。2009 年からは LHC の運転が開始され検出器からのデータが流れ始めた。現在では世界中の 400 を超える研究機関が EGI と OSG に組織されており、WLCG として 30 万を超える CPU コア、180PB を超えるディスクが接続されている。CERN の Tier0 と各地の Tier1 の間、及び Tier1 同士の間は専用のネットワーク LHC-OPN(LHC Optical Private Network)で接続され、1 接続あたり 10Gbps の帯域が確保されている。Tier1-Tier2 間は通常の学術ネットワーク(NREN: National Research and Education Network)で接続されている。

アトラス実験は WLCG の最も多くの資源を活用している。世界の 36 カ国 100 を超える研究機関がアトラス実験仮想組織に資源を提供しており、グリッド上のデータ総量で 115PB、ファイル数で 3.5 億に達する。また、再構成プロセスやモンテカルロシミュレーション、ユーザ解析など常時 15 万に及ぶジョブが実行されている。

4. 東京大学アトラス実験データ地域解析センター

アトラス実験は巨大な国際共同研究として進められている。世界 37 カ国の 170 研究機関から 3,000 人を超える研究者・大学院生が参加している。日本からも 16 研究機関、110 名が実験に従事している。参加国の研究グループはそれぞれの特長を生かして実験に貢献している。その中でデータ解析についてはそれぞれの所属研究者に応じて計算資源を提供することが求められている。日本の場合、東京大学がデータ解析のためのセンターとしての機能を担っている。

東京大学素粒子物理国際研究センターにはアトラス実験データ地域解析センター[14]が置かれている。この地域解析センターは WLCG のためのものの他に、日本の研究者が占有できる計算資源も提供している。2012 年段階では約 5,000CPU コア、約 5PB のディスクを有する。そのうち、2,500CPU コア、1.2PB のディスクを WLCG に提供している。この部分の資源は gLite ミドルウェアにより、世界中のアトラス共同研究者が利用している。システムの稼働率は年間を通した平均で 98%を超えており、安定な運用が続いている。その結果世界中の研究者が好んで使うサイトになり、ユーザ解析ジョブの実行数では Tier1 を含めて世界で 8 位にランクされている。

このように東大の地域解析センターが非常によく利用されている背景には国際ネットワークの安定性がある。東京と欧州の間のネットワーク接続は情報学研究所の運用する SINET4[15]に全面的に依存している。SINET4 は米国西海岸(Pacific Wave)まで 10Gbps、東海岸(Manhattan Landing)まで 10Gbps の帯域を持つ。

欧州にはさらに東海岸から欧州学術ネットワーク Geant[16]を経由して接続される。東京の地域解析センターは Tier2 として、運用上フランス・リヨンの IN2P3 計算センターと強く結びついている。リヨンへは Geant2 からフランス学術教育ネットワーク Renater を経由して接続されるが、東京大学より IN2P3 計算センターまで通して 10Gbps 以上の帯域が確保されている。

長距離で効率よくデータ転送を行うには特別な配慮が必要である。実際、東京からリヨンでは 300 ミリ秒近いラウンドトリップタイム(RTT)がかかっている。一定の転送の後に必ずハンドシェイクが行われる TCP 転送では効率的な転送は期待できない。実際単一接続では数 100kB/秒の転送が限界である。そこで、ファイル転送を行う場合には数十本の TCP コネクションを同時に張り、平行して転送を行う。また、たくさんのサーバを双方に準備し、複数のファイル転送を同時に行う。こういった方法により、国際ネットワークが安定なときには欧州から 500MB/秒の持続的転送を実現している。このことから、地球の裏側にあっても、LHC 実験データ解析網の一翼を十分に担えることが証明された。

実際には様々な問題がある。たとえば TCP 接続の場合、いったんパケットロスが発生すると転送効率は一気に落ちる。これは RTT が大きければ大きいほど顕著である。また、転送経路が非常に長いと言うことは、その経路のどこかで錯綜が起こる確率が高くなるということである。錯綜は制御できるものではないが、ネットワークの混雑状況を常時監視し、より安定な経路を確保することが非常に重要になる。SINETをはじめ、ネットワーク提供者との緊密な連携が望まれる。

5. 最初の三年の経験と教訓

LHC 加速器は 2010 年に本格運転を開始した。2010 年と 2011 年は加速エネルギー 3.5TeV で、2012 年は 4TeV で運転し、アトラスは 2010 年に 40pb⁻¹、2011 年に 4.7fb⁻¹ のデータを蓄積した。2012 年には 9 月段階ですでに 14fb⁻¹ を蓄積している。これらの数字は積分ルミノシティと呼ばれ、どのくらいの量の衝突をしたかを表している。

アトラス検出器は当初予測を上回る量のデータを記録し続けている。平均して 200Hz で事象データを収集するのが当初の設計であったが、実際にはビーム強度の高い衝突開始時には 1kHz、全体を平均しても 400Hz でデータを記録している。このデータ増加分はデータ圧縮などの技法を用いてカバーしている。

1 つのビームバンチは数千億個の陽子で出来ている。加速器でビームが細く絞られているため、1 回のビームバンチの交差でたくさんの陽子陽子衝突がバンチ内のそこかしこで起こる。平均発生個数は 40 近くに及んでいる。このことは事象データに含まれる粒子の数が非常に大きくなることを意味する。その結果、事象データの再構成にかかる CPU タイムが長くなる。同じ状況をモンテカルロシミュレーションで再現しようとするところからもより多くの CPU タイムが必要になる。このように計算資源に対する要請も当初計画より厳しくなっている。

このような状況の中でもデータは順調に蓄積・解析されてきた。すでに述べたように、グリッド上に蓄積されたデータ総量は 100PB を超え、ファイルの数も数億に及んでいる。これはデータ管理機能としては空前の容量である。グリッド技術がそれを十分に実現していることは注目値する。これらのファイルは同じ内容のもの複製(レプリカ)も含んでいる。より効率的にデータ解析を行うためには、計算資源(CPU)と記録資源(ディスク)を効果的に結びつける必要がある。短期間に解析を完了するために、ファイルのレプリカを分散配置し、ユーザはそのファイルの近くにある、空いている計算資源を活用して解析プログラムを走らせる。アクセスが集中するファイルはより多くのレプリカを作ることで負荷分散を図る。

複数のレプリカを用意する意味はそれだけではない。計算システムは時として障害が発生する。その内容によってはファイルが消失する場合もある。そのような場合も、複数のレプリカが存在することにより、データそのものがなくなることはなく、システム復旧後直ちにコピーを作れる。危機管理の観点から大変好ましい。検出器から取り出された生データは最も重要であり、CERN も含め、世界中に少なくとも 3 個のレプリカが確

保されている。

もう一つ重要な側面がある。それは、研究者が世界に分散しており、その誰もが平等にデータ解析を行えるシステムになっていることである。世界中のどこにいても最新のデータに機動的にアクセスできることがこれによって保証される。ユーザは自分の端末から解析したいデータセット(ファイルの集合体)を指定すると、彼のジョブが指定されたデータセットを保持するいくつかのサイトのうち、最も CPU に余裕のある場所に送られる。日本のように実験現地から遠く離れたところにいる研究者も不利になることはない。国際共同研究として倫理的に重要な側面である。

アトラスに割り当てられたジョブスロット(CPUコアに相当)は15万ほどあり、年間を通して常時再構成やシミュレーション、解析などのジョブが走っている。このように WLCG 上でこれまでに蓄積されたデータを解析が行われた結果、ヒッグス粒子と考えると差し支えない新しい粒子が発見された。これは取得された数十億に及ぶ事象の中から、崩壊モード(新粒子が崩壊して既知の粒子のグループに変換する、その組み合わせ)によっては数個から数十個程度を選び出す作業になる。7月4日の記者発表時には6月中旬まで収集されたデータの解析結果が報告された。わずか2週間の間に解析が成し遂げられたことになる。これはコンピューティンググリッド技術無しには成し得ない。

6. まとめ

LHC 加速器は順調に運転を続け、アトラスと CMS という2つの汎用検出器(発生するあらゆる事象を検出することを目指す検出器)で、これまで長年探索が続けられてきたヒッグス粒子と思われる新粒子が発見された。これには最高エネルギーをもたらす加速器技術、最先端技術の粋を集めた検出器技術によっているのはもちろんであるが、世界中に分散した計算資源をネットワークでつなぎ、統一的に運用するコンピューティンググリッド技術が重要な役割を果たしていることがわかる。

LHC 加速器は年末まで実験を継続し、昨年取得されたデータの4~5倍を今年中に集積する見通しである。その後、2013年と2014年は運転を停止し、加速器は設計エネルギーである7TeVで運用するための様々な補強作業を行う。また、検出器側も改良や補修をその間に行い、2015年に実験を再開する。その後も加速器の強化や検出器の更新を行い、最終的には2030年まで、これまで得られたデータの200倍を蓄積する予定である。

ビームエネルギーやビーム強度が高まり、また、検出器更新により読み出しチャンネル数が増加していくことから計算資源に対する要請はいよいよ高くなっていく。それに応えるための研究開発も持続的に行われている。たとえばクラウドコンピューティングはすでに計算資源の一部として WLCG に組み込まれている。今後、学術用クラウドの配備が各国で予定されており、それらの有効利用も視野にある。国際ネットワークの安定かつ効率的利用は最重要課題であるが、品質保証や他プロジェクトとのフェアシェアを実現するためのネットワークモデルである LHCONe[18]の研究も進んでいる。計算資源の仮想化技術も重要である。WLCG のサービスの配備や、ユーザ解析プログラムの移植等も仮想化により大幅に簡素化される。マルチコア・メニーコアの CPU 利用も重要な課題である。我々のアプリケーションは高い入出力帯域を要求するし、検出器の幾何学情報など莫大な仮想記憶を消費する。こういったプログラムをいかに効率的にマルチコア CPU にマップするかの研究も進んでいる。

ヒッグス粒子の発見は LHC 実験の課題の一つであったがすべてではない。これからヒッグス粒子の詳細な性質が明らかになるが、そこから、現在の我々の理解である標準模型理論の限界(高い精度で成り立っているが、いくつかの重要な疑問に答えていない不完全な理論)を超えるための手がかりが得られる。また、LHC での新粒子探索は始まったばかりで、非常に軽い質量領域が調べられたに過ぎない。素粒子物理の新しい局面が今開かれたばかりである。これからの展開を楽しみにしていただきたい。IT 技術の進展が素粒子物理の発展に果たす役割の大きさを感じつつ。

[参考文献]

- [1] CERN European Organization for Nuclear Research, <http://public.web.cern.ch/public/>
- [2] “LHC Machine”, L. Evans and P. Bryant (Editors), JINST 3 (2008) S08001.
- [3] “The ATLAS Experiment at the CERN Large Hadron Collider”, ATLAS Collaboration, JINST 3 (2008) S08003.
- [4] “The MONARC Project”, <http://monarc.web.cern.ch/MONARC/>
- [5] “The Grid: Blueprint for a New Computing Infrastructure”, I. Foster and C. Kesselman (Editors), 1999, Morgan Kaufmann Publishers. ISBN 1-55860-475-8
- [6] “Worldwide LHC Computing Grid”, <http://wlcg.web.cern.ch/>
- [7] “EGEE: Enabling Grids for E-science”, <http://www.eu-egee.org/>
- [8] “gLite - Lightweight Middleware for Grid Computing”, <http://glite.cern.ch/>
- [9] “NORDUGRID”, <http://www.nordugrid.org/>
- [10] “ARC: Advanced Resource Connector”, <http://www.nordugrid.org/arc/>
- [11] “The Open Science Grid”, <https://www.opensciencegrid.org/bin/view>
- [12] “European Grid Initiative”, <http://www.eji.eu/>
- [13] “European Middleware Initiative”, <http://www.eu-emi.eu/>
- [14] 東京大学アトラス地域解析センター計算機システム, <http://www.icepp.jp/>
- [15] 学術情報ネットワーク[サイネット・フォー], <http://www.sinet.ad.jp/>
- [16] “GEANT Project Home”, <http://www.geant.net/pages/home.aspx>
- [17] “GIP RENATER”, <http://www.renater.fr/?lang=en>
- [18] “LHCONE: LHC Open Network Environment”, <http://lhcone.net/>

ペタバイトのデータの海に
ヒッグス粒子を追う
—LHCアトラス実験の物理解析—

東京大学素粒子物理国際研究センター
坂本 宏



CERNセミナー
「Latest update in
the search for the
Higgs boson」
2012年7月4日

©CERN

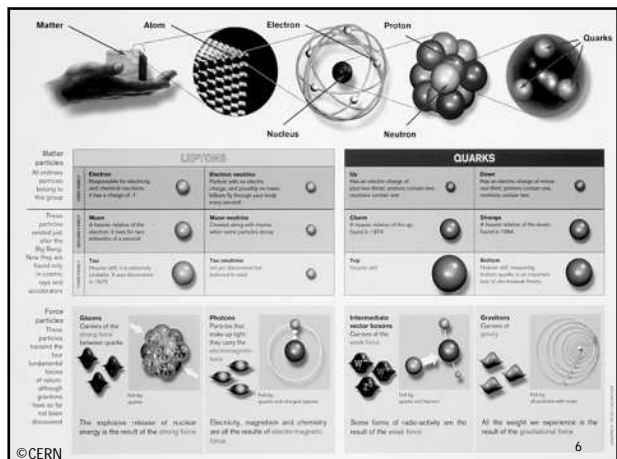


目次

- 現在の素粒子像
- LHC加速器とアトラス実験
- データ解析への要請
- 世界LHC計算グリッド
- 東京大学地域解析センター
- データ解析の現状

現在の素粒子像

- 素粒子の標準模型
 - 宇宙を構成する粒子 フェルミオン
 - 力を伝える粒子 ベクトルボゾン
 - 質量の獲得 ヒッグス場
- 力の統一
 - 4つの力
 - より高いエネルギーでは同じ振る舞い
 - ビッグバンに近づく



UP QUARK CHARM QUARK TOP QUARK
DOWN QUARK STRANGE QUARK BOTTOM QUARK
PHOTON GLUON
ELECTRON-NEUTRINO MUON-NEUTRINO TAU-NEUTRINO
ELECTRON MUON TAU
W BOSON Z BOSON HIGGS BOSON

<http://www.particlezoo.net/> 7

ドブロイ波長

$$\lambda = \frac{h}{mv} = \frac{h}{p} = \frac{2\pi}{k}$$

加速電圧	ドブロイ波長
100V	0.12nm(10 ⁻⁹ m)
1kV(10 ³ V)	0.03nm
1MV(10 ⁶ V)	0.84pm(10 ⁻¹² m)
1GV(10 ⁹ V)	1.2fm(10 ⁻¹⁵ m)
1TV(10 ¹² V)	1.2am(10 ⁻¹⁸ m)

Instruments: Accelerators (LHC, LEP), Electron Microscope, Microscope, Telescope, Radio Telescope

Observables: SUSY particle?, Higgs?, Z/W, Proton, Atom, Virus Cell, Earth radius, Earth to Sun, Galaxies, Radius of observable Universe

Big Bang

8

出典: JEOL(日本電子株式会社)

K Si Fe 1.0nm

9

History of the Universe

Accelerators: CERN-LHC, Fermilab, SLAC, DESY, KEK, J-PARC, SNS, JLAB, MAMI, THALES, DESY, KEK, J-PARC, SNS, JLAB, MAMI, THALES

Key: quark, gluon, electron, photon, meson, baryon, star, galaxy, black hole, neutrino, atom

Particle Data Group, LBNL, © 2000. Supported by DOE and NSF

10⁻³⁴ sec 10⁻³² sec 10⁻¹⁶ sec 10⁻¹⁶ sec 100 sec 10000 years

10⁻¹⁶ m 10⁻¹⁶ GeV 10⁻¹⁶ m 10⁻¹⁶ GeV 10⁻¹⁶ m 1 MeV 10⁻¹⁶ m 10 eV

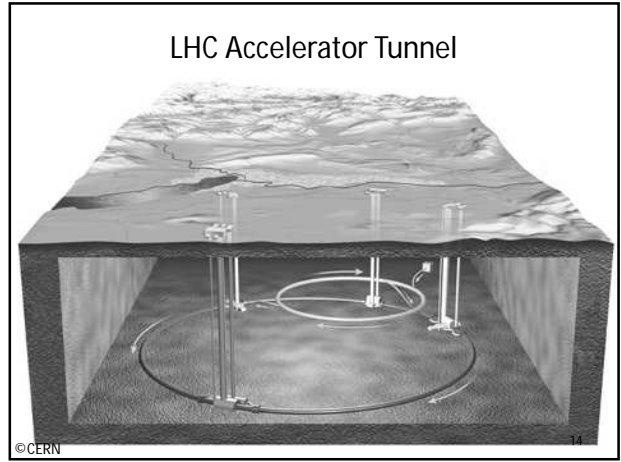
Quantum Gravity
Super Unification
Grand Unification
Electroweak Model
QED Electro Magnetism
Weak Theory
QCD
Universal Gravitation
Gravity

Magnetism
Electricity
Weak Force
Nuclear Force
Celestial
Terrestrial
Gravity

LHC加速器とアトラス実験

- CERN(欧州原子核合同研究機構)
 - スイス・ジュネーブ郊外
 - 日本・アメリカ・カナダなども協力
- LHC加速器
 - 世界最高エネルギー陽子陽子衝突型加速器
 - 設計加速エネルギー7TeV(テラエレクトロンボルト)
 - ・ 単3乾電池を並べると太陽までの1.5倍の距離
- アトラス実験とCMS実験
 - 汎用測定器: 全周を覆い、すべての粒子を検出
 - 未知の新粒子の探索、標準理論の精密検証

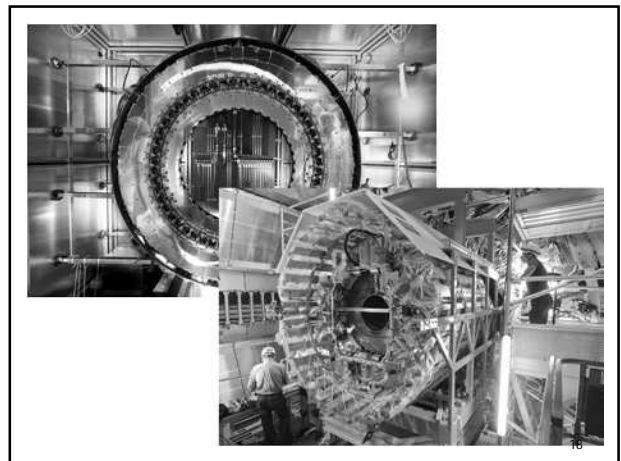
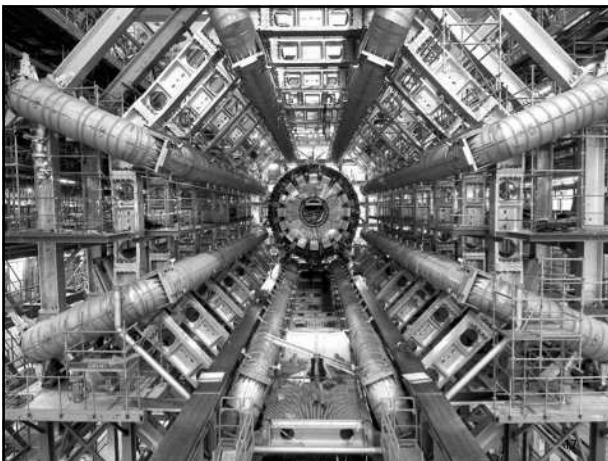
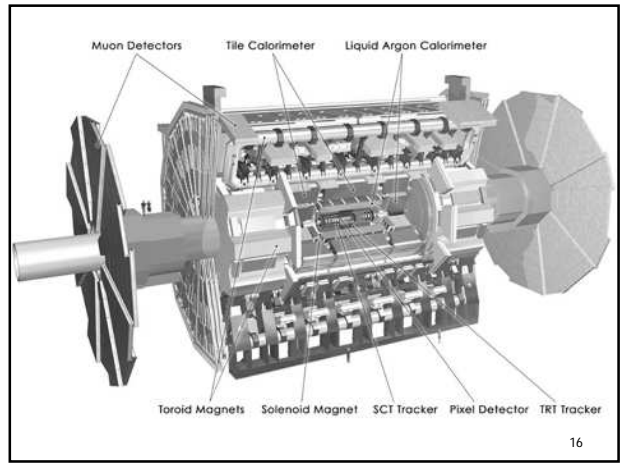
12

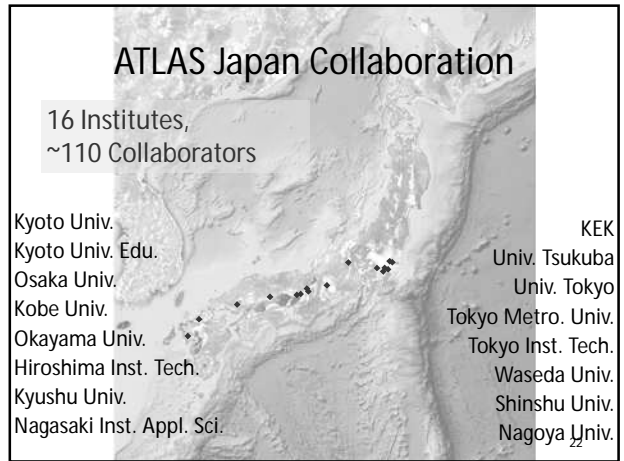
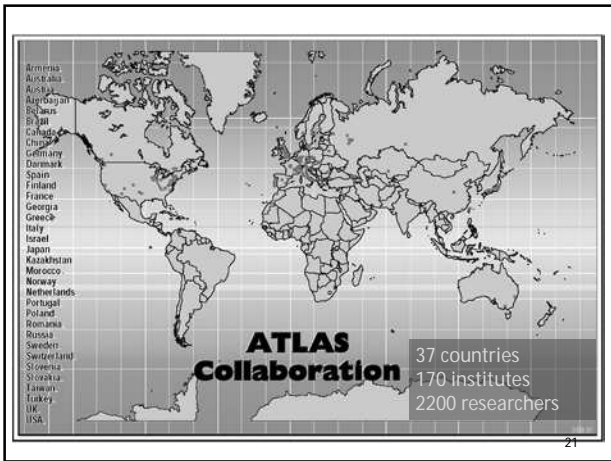
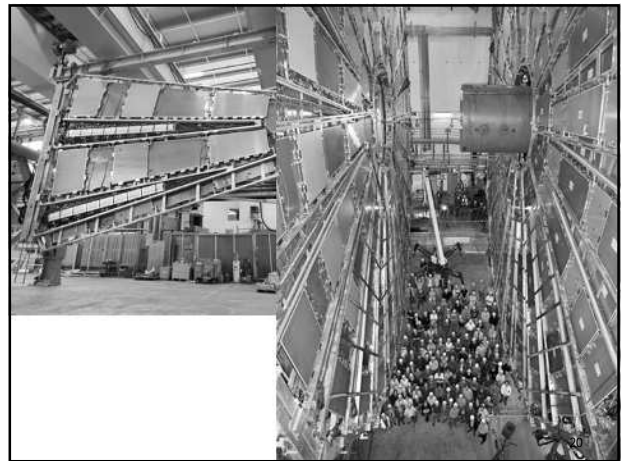
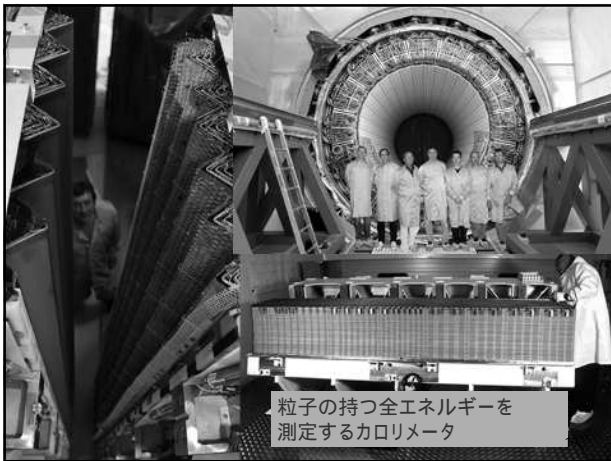


Accelerator Components

- 1232 Main Dipole Magnets
- 386 Main Quadrupole Magnets
- A few thousands of Correction Magnets.

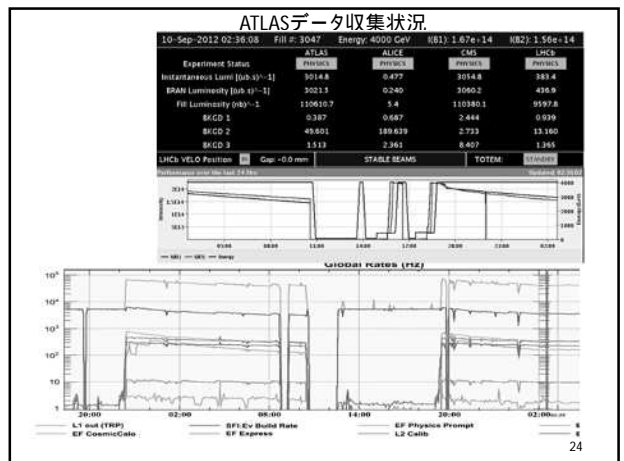
©CERN 15

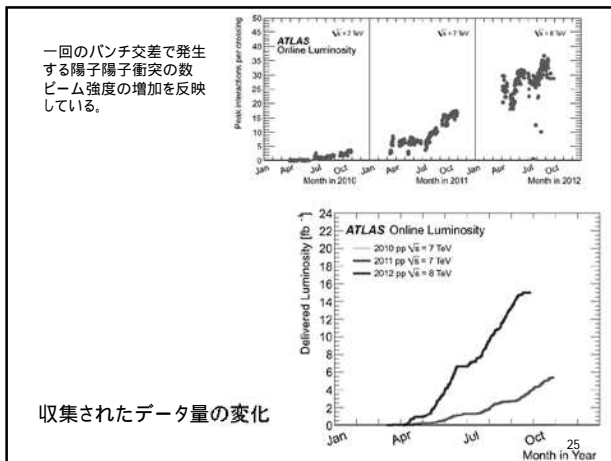




データ解析への要請

- 毎秒500MBの生データ
 - 9秒でDVD1枚分
 - 物理解析用データまで4段階の処理
- 数十万台のPCの処理能力
 - 実データの再構成と解析(1事象100秒/PC)
 - モンテカルロシミュレーションとの比較(1事象1,000秒/PC)
- 世界中の研究者がアクセス
 - 誰もが平等に解析できる~効率よく参照
 - 複数のレプリカ~安全な保管

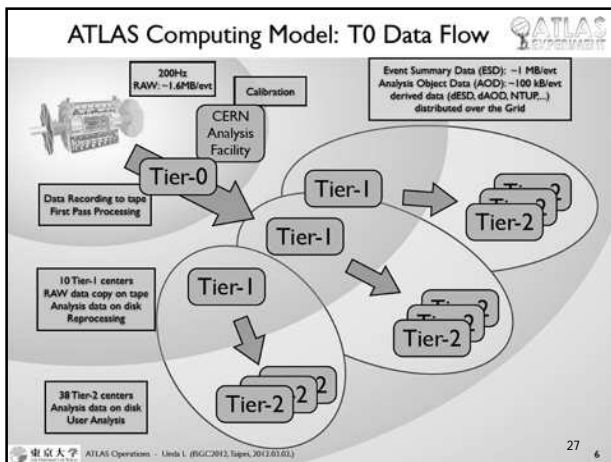




ATLAS offline computing

- Prompt data reconstruction at Tier-0 has kept up with LHC performance in last weeks
 - Extra resources used – up to 7.5k CPU cores
 - 15-40 seconds/event for RAW→ESD step
 - In line with expectations from simulation
- World-wide Grid computing is essential
 - Up to 150k simultaneous jobs running in parallel
 - Mainly MC simulation and user analysis jobs
 - Computing centres give resources beyond pledged commitments
 - Very beneficial for the broad physics program

9th July 2012 Richard Hawking 5



世界LHC計算グリッド

- 世界中の研究機関・大学を接続
 - 150に及ぶ機関の計算資源
 - 学術・教育用ネットワークによる接続
- 共通のミドルウェアの配備
 - コンピューティンググリッド技術
 - 世界で最初の学術用実用グリッド

28

なぜ今更グリッド？

- 世界中の研究機関で計算資源を共有
 - 各国は事情に応じて相応の計算資源を供出
 - 世界中の研究者が平等に資源にアクセス
- 必要な機能
 - データは核心的: 単一のデータ管理機構
 - 共通のソフトウェア: ヘテロな環境での実行
 - 効率的な運用: 広域ネットワークへの要請
 - 安全なアクセス: 認証基盤と仮想組織

29

合同分科会 ～プログラム～

高速ネットワークを考える

1. 目的 2011年10月11日(水) 13:00～21:30
11月1日(水) 9:00～20:30
(11月2日(木) 9:00～14:00, 研究センターオンライン)

2. 場所 特別イベントシステムホストルーム 3階 (六本木) 申込 速見氏

3. プログラム 合同分科会(CM01)主催スラング

14:00～11:00 懇話会

合同分科会2003～プログラム～

グリッドコンピューティングを考える

1. 日時 2003年10月22日(水) 13:00～21:00
2003年10月23日(木) 9:00～21:30
2003年10月24日(金) 9:00～13:30

2. 場所 特別イベントシステムホストルーム 3階 (真珠の国) 申込 速見氏

3. プログラム (速見氏)

12:30～13:00 受付 申込 (後援社) 明治大学

13:00～13:30 速見氏 (伊豆山研究所) 講演

13:30～13:35 速見氏 (理研) 講演 (理研) 講演 (理研) 講演 (理研) 講演

13:35～14:00 特別講演 (What's GRID?) 早稲田大学 速見氏 共同講演

14:00～14:30 GRID Forumにおける最新の議論に加えて、速見氏はLHCのグリッド環境の構築について、最新の情報を提供し、議論を促す。また、速見氏は、グリッドという言葉を理解しているもの、または興味を持っている方に対しては、LHCのグリッド環境があると思います。この講演では、これらも踏まえて「GRID」の話をさせていただきます。

30

isgtw international science grid this week

2001.1~2004.3

2004.4~2010.1

2010.2~?

Happy 10th Birthday, WLCG!

calendar

Independent experimental data challenges

Service challenges proposed in 2004 for the intermediate service experts

- Data transfers for weeks on-end
- Data management
- Scaling of job workloads
- Security incidents ("free data")
- Interoperability
- Support processes

2004

2005

2006

2007

2008

2009

2010

Comics-only data

LHC data

31

The Grid: Blueprint for a New Computing Infrastructure
I. Foster and C. Kesselman (Eds), Morgan Kaufmann, 1999.

32

グリッドの構成要素

- 分散ジョブ～計算資源管理
 - ワークロード管理～負荷分散
 - グリッドジョブ処理～ジョブ記述・キュー
- データ共有～記憶資源管理
 - ファイル転送
 - レプリカ管理
- 情報システム
 - 資源情報の発信
 - サイトモニター
- 認証基盤
 - 仮想組織
 - アクセス制御

33

WLCG (Worldwide LHC Computing Grid)

- LHC4実験のコラボレーション組織
 - 参加国 参加研究機関
- LHC実験データ解析用グリッドの配備
 - OSG(米国), Nordugrid(北欧)とgLite(他地域)
- 相互運用可能
 - 各実験はそれぞれ上位階層を用意

34

Running jobs: 226115
Transfer rate: 12.00 GiB/sec

35

EGI Resource Centres (April 2012)

5/22/2012 Slide borrowed from Steven Newhouse

36

EGI Metrics (April 2012)

Metrics		Value (Yearly increase)
Installed Capacity	Logical CPUs	270,800 (+30.7%)
	EGI-InSPIRE and EGI Council members	399,300
	Including integrated RPs	2.96 Million (+49.5%)
	HEP.SPEC 06	2.96 Million (+49.5%)
Storage	Disk (PB)	139 PB (+31.4%)
	Tape (PB)	134.3 PB (+50%)
Resource Centres	EGI-InSPIRE and EGI Council members	326
	Including integrated RPs	352
	Supporting MPI	90
Countries	EGI-InSPIRE and Council members	42
	Including integrated RPs	54
Performance	Monthly Availability/Reliability	94.50%/95.42%
Utilization	HEP.SPEC 06 Hours	10.5 Billion (+52.91%)
	Jobs	492.5 Million Jobs /year 1.35 Million Jobs/day (+46.42%)

5/22/2012

Slide borrowed from Steven Newhouse

37

OSG resources

- Resources accessible through the OSG are contributed by the community
 - Their autonomy is retained.
 - Resources can be distributed locally as a campus infrastructure
- >100 sites
- >70,000 cores accessible
- >30 research communities



5/22/2012

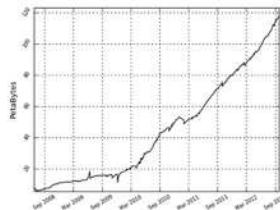
Slide borrowed from Lothar Baurdick

38

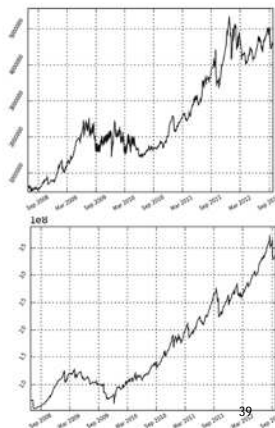
世界中に配布されたアトラスデータの総量

100PB(ペタバイト)
DVD2,000万枚に相当
数百兆回の衝突
数十億事象のサンプル

Total GRID space usage according to DQ2



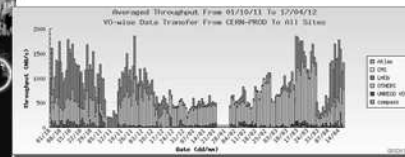
Total GRID datasets according to DQ2



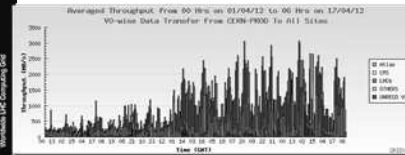
39

Data transfers...

Already back to "normal" levels for accelerator running



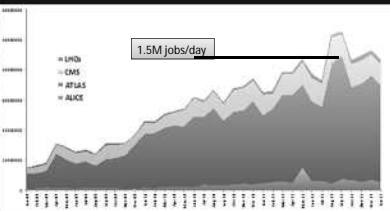
Since last RRB



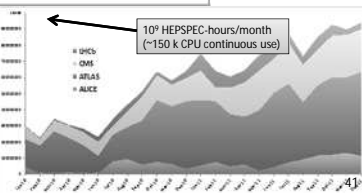
2012 Data

40

WLCG in 2011

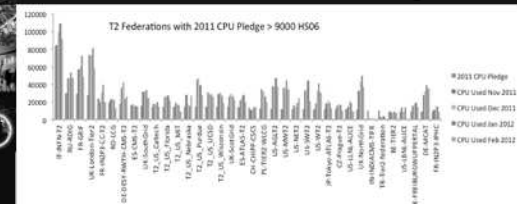


Usage continues to grow
- # jobs/day
- CPU usage

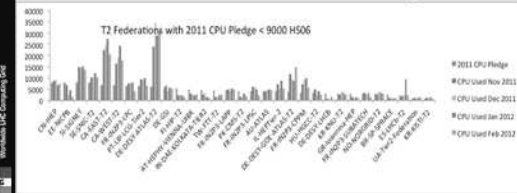


41

Tier 2 accounting



2011 CPU Pledge
CPU Used Nov 2011
CPU Used Dec 2011
CPU Used Jan 2012
CPU Used Feb 2012



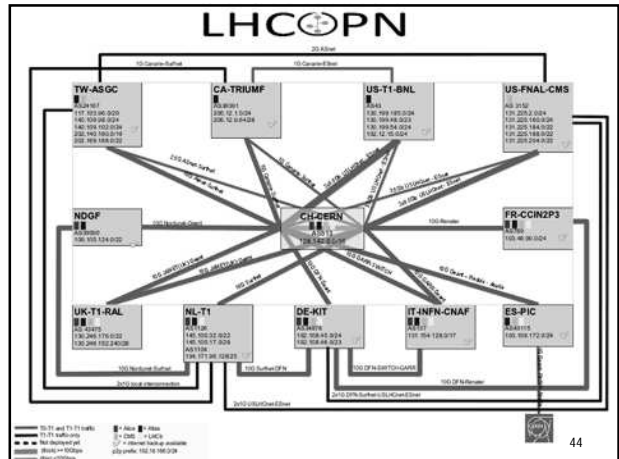
2011 CPU Pledge
CPU Used Nov 2011
CPU Used Dec 2011
CPU Used Jan 2012
CPU Used Feb 2012

42

国際ネットワーク

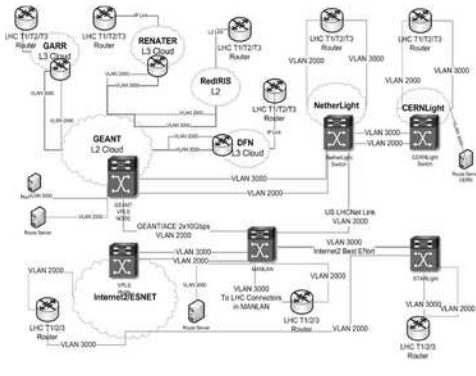
- 階層構造の実現(当初モデル)
 - Tier0-Tier1/Tier1-Tier1: LHC-OPN (Optical Private Network)
 - Tier1-Tier2: NREN (National Research and Education Network) Tier1: Tier2=1:n
- ネットワークポロジィーの変化・発展
 - Tier1: Tier2=m:n
 - Tier2-Tier2
 - LHC-ONE (Open Network Environment)

43



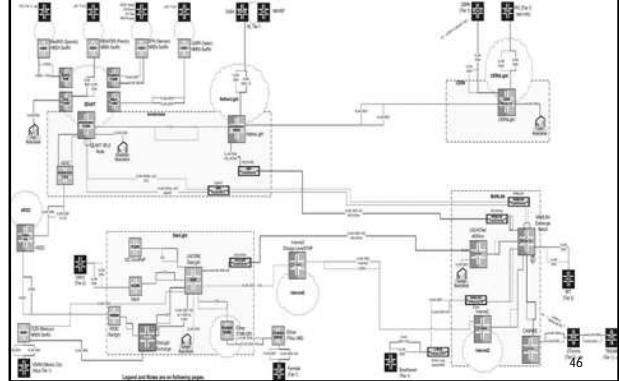
44

LHCONE Overview (from Dante)



45

LHCONE Detailed View (Bill Johnson)



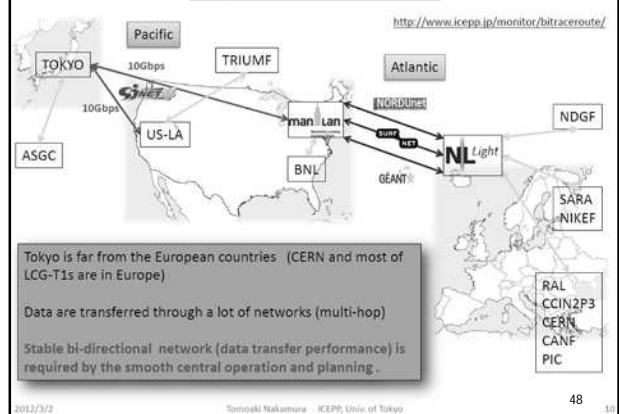
46

東京大学地域解析センター

- 日本の研究者のための解析施設
 - 東京大学素粒子物理国際研究センターに設置
- 世界LHC計算グリッドサイト
 - CERNとの覚え書きに基づいて運用
 - 著者数比に応じて計算資源を提供
 - 世界中の共同研究者が利用
- CERNサテライトシステム
 - CERN現地に東京大学が設置
 - 現地滞在の日本の研究者専用

47

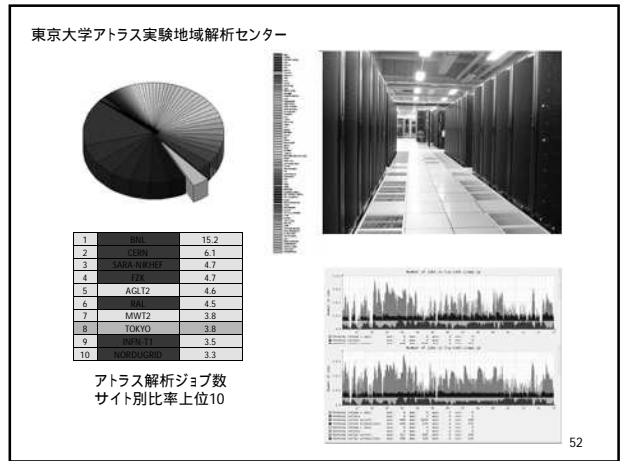
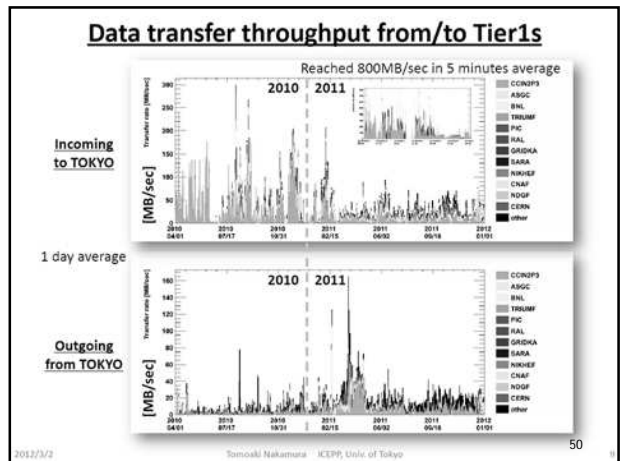
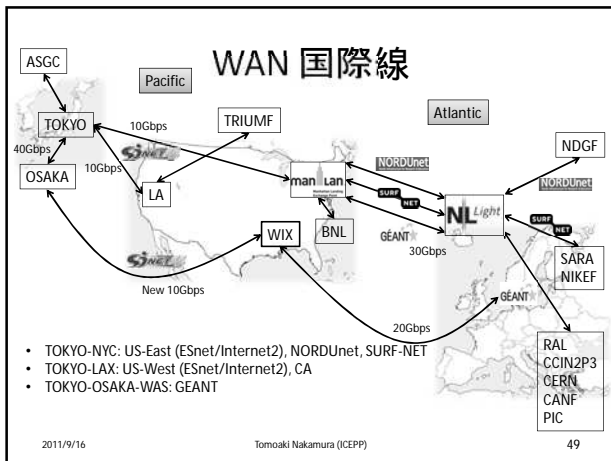
Data transfer via WAN



2012/3/2

Samuaki Makemura ICEPP, Univ. of Tokyo

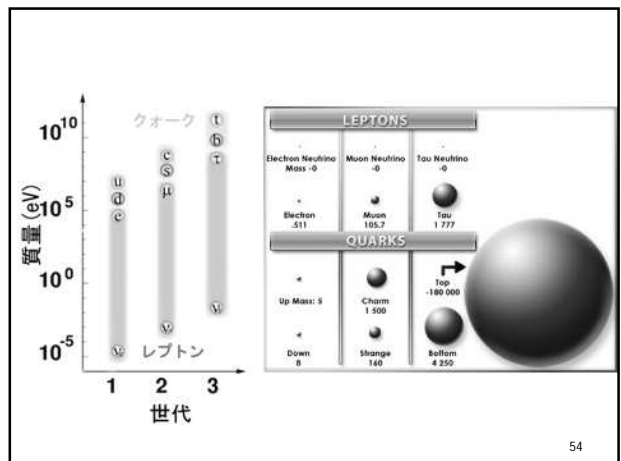
48



データ解析の現状

- 2010年に衝突実験を開始
 - 2011年衝突エネルギー7TeV
 - 2012年前半、衝突エネルギー8TeV
- 質量126GeV付近に新粒子発見
 - 光子やZ粒子などに崩壊
 - ヒッグス粒子と考えられる

53





©CERN

55



©CERN

56



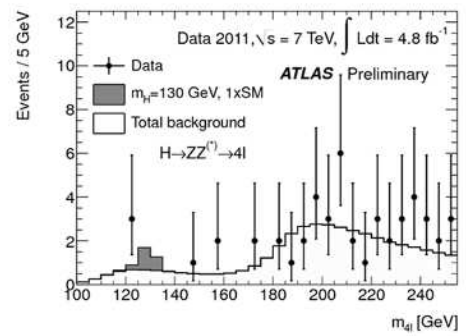
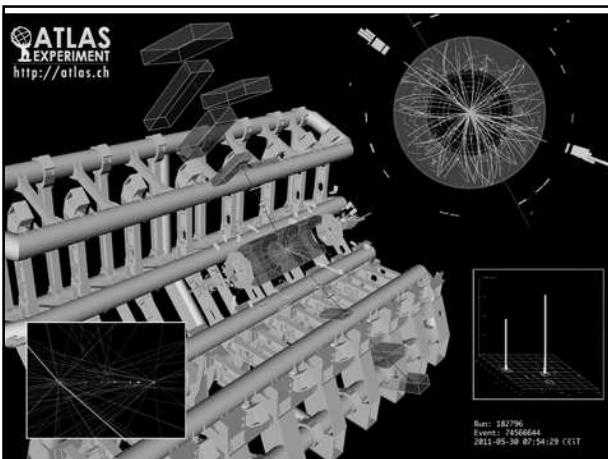
©CERN

57

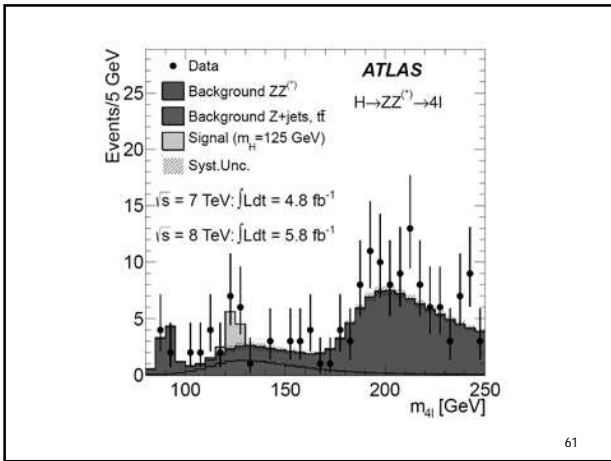


©CERN

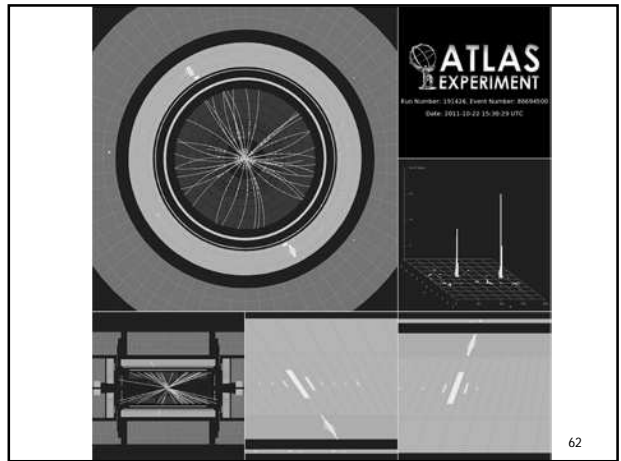
58



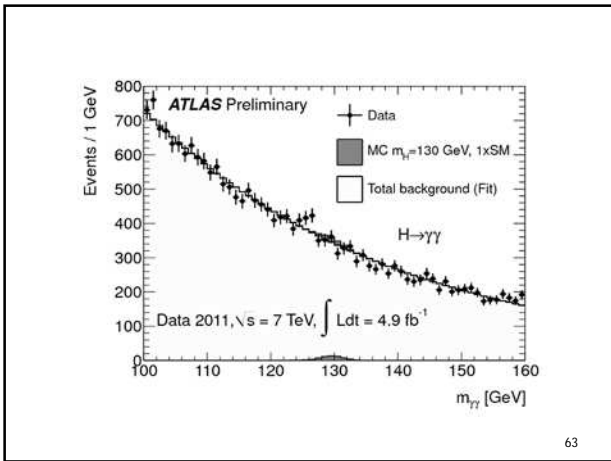
60



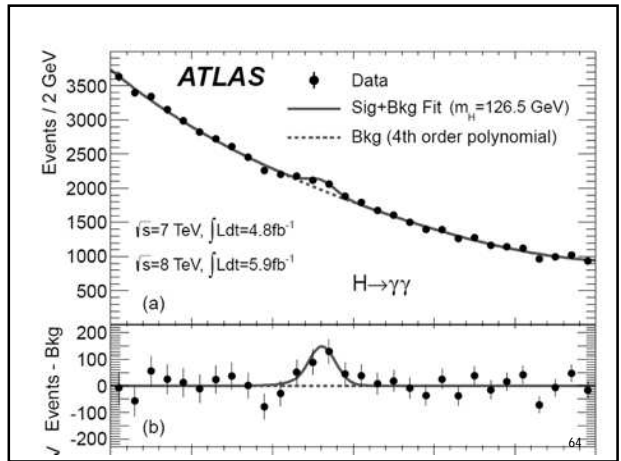
61



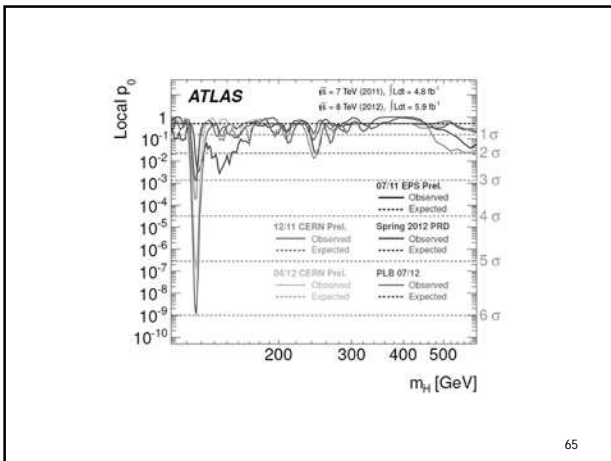
62



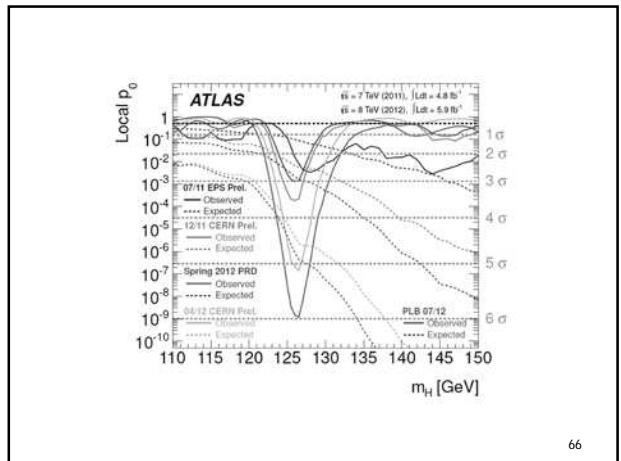
63



64



65



66

今後の計画

- 2012年末まで8TeVでの衝突実験継続
 - 新粒子の崩壊を様々なモードで探索
 - 新粒子の性質を精査～ヒッグス粒子か？
- 2013年～2014年加速器調整のため停止
 - 2015年から衝突エネルギー14TeVで実験
 - より高いビーム強度での実験
 - 新しい素粒子の世界に踏み込む

67

アップグレード

- クラウドコンピューティング
 - グリッドのCEとして取り込む
- 仮想化技術
 - ヘテロ環境での差違の吸収
- マルチコア
 - アプリケーションソフトウェアの再構築
- ネットワークQoS
 - 専用帯域の確保・他プロジェクトとの共存

68

まとめ

- 世界最高エネルギー加速器LHC
 - 高いエネルギーと衝突頻度を実現
- アトラス実験
 - 1億チャンネルのセンサーで事象を完全に捕獲
- WLCG世界解析網
 - 1億分の1の頻度の事象を選び出す
- 加速器技術、検出器技術、IT技術の総合
 - すべてがそろって初めて実現

69

合	同	分	科	会	選	出
---	---	---	---	---	---	---

合同分科会 2012 年度会合 より

福島原発事故

- 情報発信とデータのアーカイブ -

早野 龍五

(東京大学)

福島原発事故

情報発信とデータのアーカイブ

東京大学 早野 龍五



twitter hayano

ビッグデータは 人を救えるか？

Part 1 - 福島原発事故と僕 - why am I here?

震災前の僕

妹ヨハルシケンキユ

なぜ宇宙にはないのか

陽電子 (電荷+1)
反陽子 (電荷-1)
反水素原子 (原子番号 -1)

電子 (電荷-1)
陽子 (電荷+1)
水素原子 (原子番号 1)

反物質

物質

宇宙で最も多い元素

FACT X-33 CERN-1 CERN-2 BUILDING C LHC ANTIMATTER G... God

物理学者とともに読む
「天使と悪魔」の虚と実 50のポイント

東京大学教授 早野龍五
CERN研究所における
反物質研究グループのリーダー
<http://nucl.phys.k.u-tokyo.ac.jp/hayano/>

本サイト公開後の対応はこちら。

「天使と悪魔」 上巻に登場する主な科学者
ロバート・ラングドン
ハーヴァード大学教授 同数論学専門
マクシミリアン・コーラー
セルジ (粒子線子核研究機構) 所長
レオナルド・ヴェトラ
セルジの科学者 カトリック司教
ヴィットリア・ヴェトラ
セルジの科学者 レオナルドの妻

In collaboration with
角川文庫 14276, 14277, 14278 帯掘
ソニーピクチャーズ 2009.5.15封切

上巻の巻末には、CERN研究所が反物質の生成に成功し、新型の反陽子減速装置の開発に着手

1. 「スイスのセルンが、反物質粒子の生成に先を初めて成功した」
これは、1996年の「Physics Letters」誌に、CERNで実験していた

1.専門は物理実験

ジュネーブのCERN研究所で「反物質」研究

2.新指導要領の中学理科教科書

平成20年3月 平成22年11月一部改正 (中学校 2012年4月~施行)

文部科学大臣 渡海 紀三朗

新学習指導要領・生きる力

理科 [第1分野]

原子の成り立ちについては、原子が電子と原子核からできていることを扱うこと。その際、原子核が陽子と中性子でできていることにも触れること。放射線の性質と利用にも触れること。

理科 [第2分 30年ぶりの復活

... (7) 科学技術と人間

ア エネルギー

(イ) エネルギー資源

人間は、水力、火力、原子力などからエネルギーを得ていることを知るとともに、エネルギーの有効な利用が大切であることを認識すること。



岩波「科学」10月号

市民科学と放射線教育 東大教育学研究科 小玉重夫氏

(中学の理科の教科書の中で)

唯一、福島第一原発の事故について言及

人体に健康被害が出るおそれを指摘

達成できる範囲で線量を低く保つようにと
うICRP勧告を取り上げ

エネルギー政策をめぐる論点の存在も指摘

1.専門は物理実験

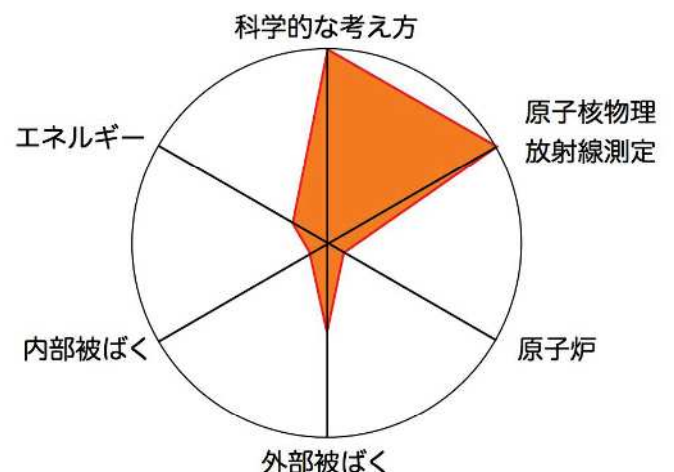
ジュネーブのCERN研究所で「反物質」研究

2.新指導要領の中学理科教科書

3.積算で200mSv被ばく

肺ガンで右肺上葉切除

2011/3/11以前の私のリテラシー



2011.3.11 14時46分
東北地方太平洋沖地震

震災直後の僕



我が家のテレビも落下



twitter havano hayano

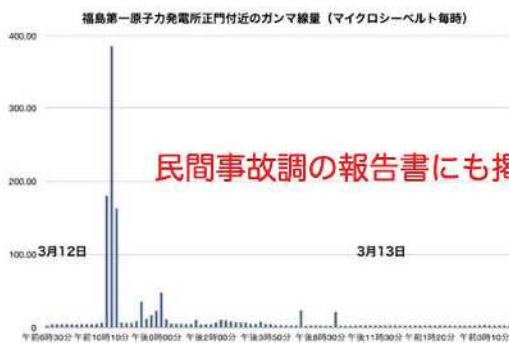
3月12日 14:22

Cs137が出す662 keVのガンマ線を
確認したという意味か。福島第一原
子力発電所。Cs137は天然には存在
せず、Sr90とともにウランの核分裂
で生じる代表的な放射性同位元素。

twitter havano hayano

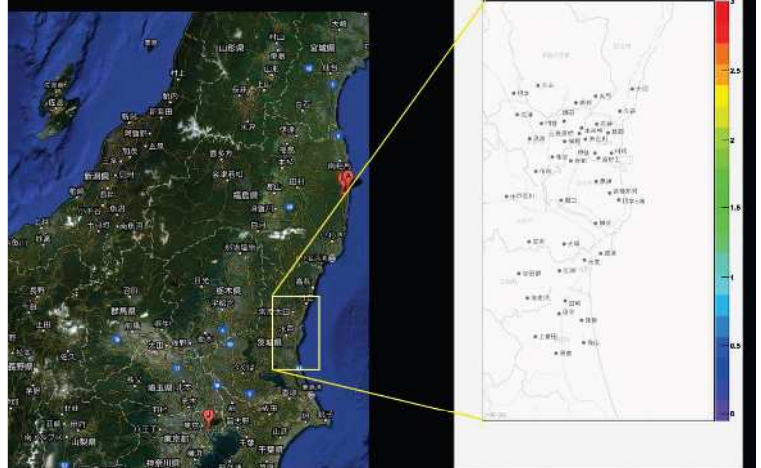
3月13日 07:49

福島第一原子力発電所正門付近のガンマ線量測定値、東電公表データ
(<http://bit.ly/engea6>) からグラフにしてみました。

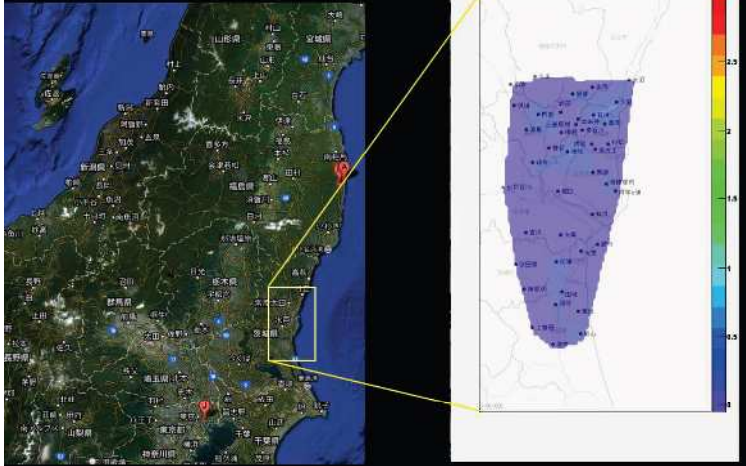


3/15朝 茨城のモニタリングポスト

raddata/2011-03-15-00:40.gif



3/21朝 茨城のモニタリングポスト



2500 (震災前) → 150,000 フォロワー

乾・岡崎 研究室

東北大学
工学部情報科学システム融合学際
コンピュータサイエンスコース
初期コンピュータインテグレーション
大学教育情報科学研究科
システム情報科学専攻
情報科学学際
情報科学学際

English / Japanese

メンバー
アクセス・連絡先
いっしょに研究会
研究内容
研究発表一覧
学位論文一覧
研究会・勉強会
研究室環境
公開資料
留学活動
アルバム
過去のニュース

ローカルページ

検索 編集 更新 名前変更 設定
バックアップ 複製 一時

FrontPage / Project 311 / トレンド分析

- ねらい
 - 組織物
 - 名詞句に対するトレンド分析
 - 名詞句の出現ペアに対するトレンド分析
 - 寄り付けの文脈ペアに対するトレンド分析
 - リツイートされたユーザーに対するトレンド分析
 - URLの普及に対するトレンド分析
- 分析手続
 - 名詞句の抽出
 - URLの抽出と短縮URLの解決
 - トレンド分析
- 更新

ねらい

ツイート上の統計情報（ツイート数やバーストなど）が、東日本大震災に関連する事象・事象をどのように反映しているのか、実証・調査しています。

成果物

名詞句に対するトレンド分析

- データの中で頻度が高い名詞句TOP100: <http://www.ci.ecei.tohoku.ac.jp/prj311/trend/noun/top100.html>
- バースト性を持つ名詞句の一覧 (時間別): <http://www.ci.ecei.tohoku.ac.jp/prj311/trend/noun/burst-time.html>
- バースト性を持つ名詞句の一覧 (スコア別): <http://www.ci.ecei.tohoku.ac.jp/prj311/trend/noun/burst-ranking.html>
- 名詞句の依存量を時系列可視化: <http://www.ci.ecei.tohoku.ac.jp/prj311/trend/trendview.py?c=noun>

Twitter Japan 株式会社より提供された179,286,297ツイートから、1,152,607,055個 (3,1206,959種類) の名詞句が抽出されました。このうち、全データにおける出現頻度が1,000回以上の名詞句 (71,691種類) を分析の対象としました。

実際にデモシステムを使っていると、東日本大震災の記憶が蘇ってきます。ぜひ各自で体験して頂きたいと思いますが、ここではいくつかのスクリーンショットを載せることにします。

リツイートされたユーザーに対するトレンド分析

- データの中でリツイートされた頻度が高いユーザーのTOP100:

<http://www.ci.ecei.tohoku.ac.jp/prj311/trend/rteduser/top100.html>

順位	ユーザ	被RT回数
1	@NHK_PR	630459
2	@nhk_seikatsu	304824
3	@Asahi_Shakai	279259
4	@nhk_news	209515
5	@nhk_HORIJUN	173995
6	@tsuda	165434
7	@hayano	145436
8	@nhk_kabun	127916
9	@earthquake_jp	114806
10	@toughokujishin	112592

twitter

の強みと限界

強み：

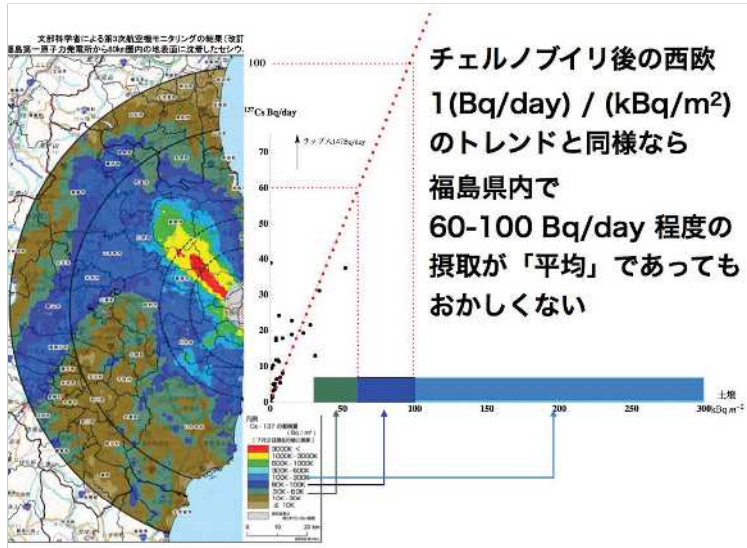
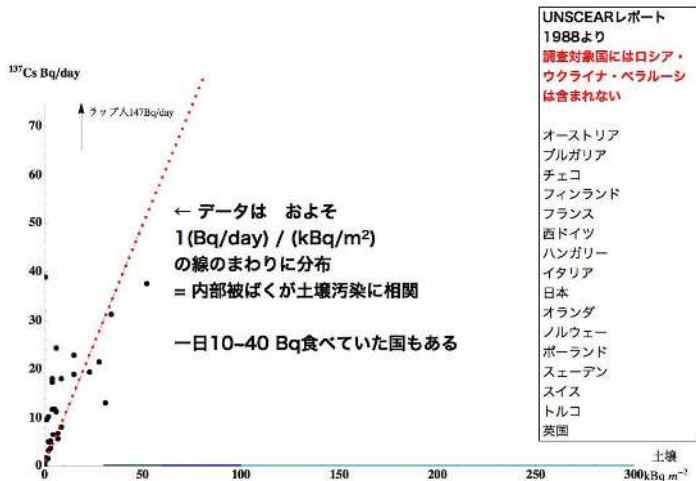
双方向

限界：

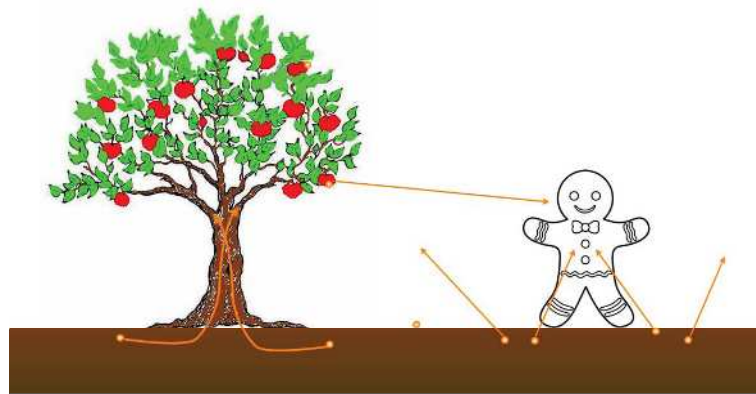
所詮“cloud”の論壇

問題解決の主戦場にはなりえない

チェルノブイリ事故後最初の1年間の土壌汚染と内部被ばくの関係
 土壌汚染と、WBCで推定した¹³⁷Csの一日平均摂取量の相関



チェルノブイリなどでは
 土壌汚染 と 外部被ばく と 内部被ばく は比例
 現在の日本ではどうか？



私はセシウムを一日あたり
 約 〇〇 Bq 食べていると思う

<1 5 10 50 100<

福島県の人
 は
 セシウムを一日あたり
 約 〇〇 Bq 食べていると思う

<1 5 10 50 100<

「給食まるごと検査」の提案
 (2011/9/21 に 森文科副大臣に提案)

三次補正予算 食品のサンプル検査	新規提案 給食一食まるごと検査
<p>簡易検査機による サンプル検査</p> <p>すり抜けても 分からない という不安</p> <p>イメージ図 2010年10月 検査の準備を 済ませた上で 検査を実施</p>	<p>子供が食べたのと同じ物を丸ごと</p> <p>ミキサーにかけて マリネリ容器に詰め</p> <p>イメージ図 2011年9月 検査の準備を 済ませた上で 検査を実施</p> <p>ゲルマニウム検出器で 精密測定</p>

南相馬市長への提案 (2011/10/28)



UNIVERSITY OF TOKYO
SCHOOL OF SCIENCE, DEPARTMENT OF PHYSICS
7-3-1 HONGO, BUNKYO-KU, TOKYO 113-0033, JAPAN

Professor Ryugo S. Hayano
早野 龍五

TELEPHONE: +81-3-5841-4235
TELEFAX: +81-3-4496-4043
e-mail: hayano@phys.s.u.tokyo.ac.jp
http://nud.phys.s.u.tokyo.ac.jp/hayano

2011年10月28日

南相馬市長
桜井勝延様
FAX 0244 24

南相馬市学校給食丸ごとセシウム検査の御提案

桜井市長、初めてお便り差し上げます、東京大学理学部物理の早野龍五と申します。
私は、子供たちが日常の食事で摂取している（かもしれない）放射性セシウムの量を把握し、内部被ばくの低減をはかるため、学校給食で実際に子供たちが食べた食事をミキサーにかけてゲルマニウム半導体検出器で精度よく測定することを各方面



紆余曲折の末 (早野の私費で) 1月より検査開始



HOME | サイトマップ | English | Chinese | Korean

HOME > 東日本大震災関係インフォメーション > 学校給食放射線量測定結果
東日本大震災関連情報

最終更新日: 平成24年2月1日

学校給食提供食の放射線量の測定結果

セシウム不検出

検査方法

(財)日本冷凍食品検査協会(横浜市金沢区福浦)に委託、ゲルマニウム半導体検出器によるガンマ線スペクトロメトリーによる核種分析
測定時間2,000秒
学校給食で実際に提供した給食1食分を1週間(5日分)ごとにまとめて測定
検体は毎日、南相馬市学校給食センターから採取

検査結果と内部被ばくの実効線量

提出期間 (日数)	合計重量 (Kg)	検査日	測定結果 (Bq/Kg)			内部被ばくの実効線量 (μSv)
			放射性 ヨウ素131	放射性 セシウム134	放射性 セシウム137	
平成24年1月23日~1月27日 (5日分)	5.0	1月30日	不検出 <0.52	不検出 <0.95	不検出 <0.57	0~0.068未満
平成24年1月16日~1月20日 (5日分)	4.9	1月21日	不検出 <0.54	不検出 <0.69	不検出 <0.88	0~0.090未満

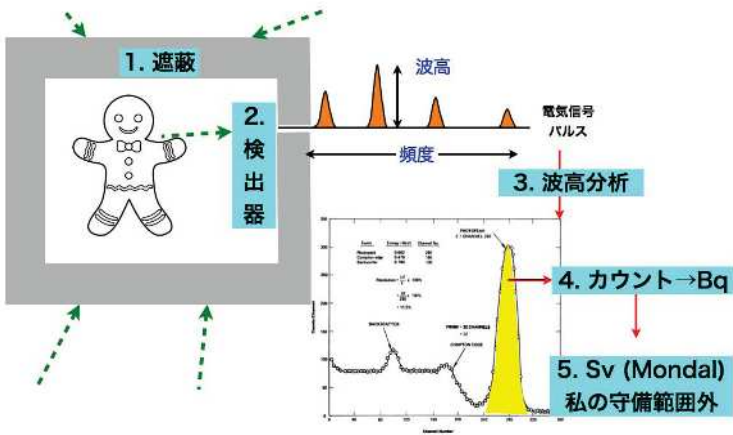


そして
WBC問題にも引きこまれた

私のtweetsを読んでおられる
福島県内のお医者さんと繋がった



装置の要素

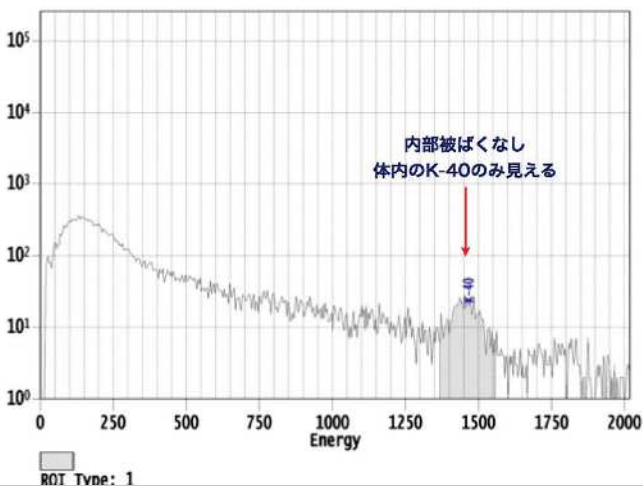


どの人にも存在する ^{40}K

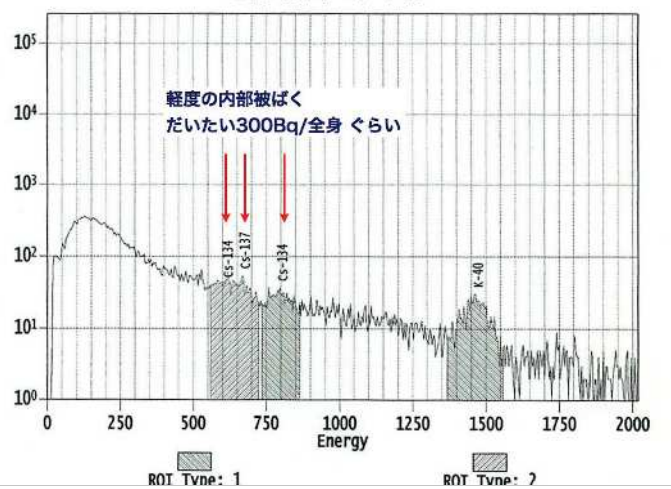


Web公開されている放医研ビデオ「【応用】立位型ホールボディカウンター」より

WBCでこんなスペクトルが見えます



ちょっと内部被ばくしている方の場合は こんなスペクトル



しかし中には

WBCで20000 BqのSさん、毎日一定量ずつの放射性セシウムを事故以来食べ続けていたと仮定して（恒常摂取）その平均摂取量を推定すると…

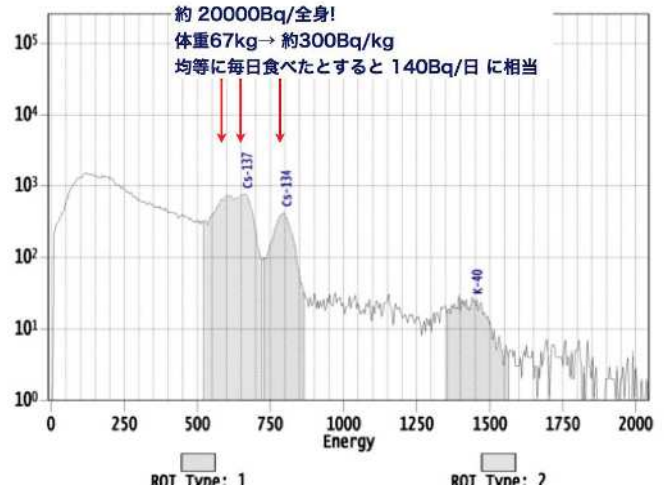
成人の場合、1Bq/日食べ続ける → 体内に約140Bq。
 $20000/140 = \text{約}140\text{Bq/日}$

田崎本p119

年齢	3ヶ月	1歳	5歳	10歳	15歳	成人
平衡量 (Bq)	23	19	30	53	117	143

表 6.3 毎日1Bqの放射性セシウムを摂取し続けた際の体内の放射性セシウムの平衡量。ICRPの動態モデルに基づいて計算した。

8/2 測定 70歳 Sさん - 男性



Sさんの年間の内部被ばく線量

$$140 \times 5.6 = 784\mu\text{Sv} \rightarrow \text{年間の実効線量は約}0.8\text{mSv}$$

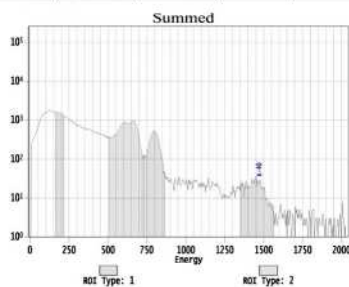
年齢	3ヶ月	1歳	5歳	10歳	15歳	成人
年間の実効線量 (μSv)	8.4	5.0	4.0	4.2	5.6	5.6

表 6.2 毎日1Bqの放射性セシウムを摂取し続けた際の年間の被曝量
 田崎本p.113より

25,000Bq/bodyのYさん

NUCLIDE RESULTS FOR FASTSCAN <Summed>

Nuclide Name	Id Confidence	Wt Mean Activity (kBq)	Error (2SD)	Action Level 1	Action Level 2	MDA (kBq)
K-40	1.000	4.265189E+000	20.06%			2.927E+000
Cs-134	0.990	9.113656E+000	4.63%	Not Performed	Not Performed	5.616E+001
Cs-137	0.996	1.591799E+001	5.84%			5.934E+001



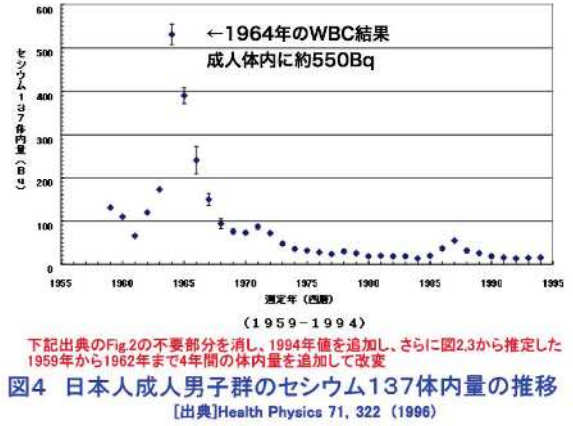
H24.9.18に葛尾村で今後除染作業をするためにWBC検査

避難指示解除準備区域の方、奥さんは避難したままだが、Yさんは自宅で現在生活している

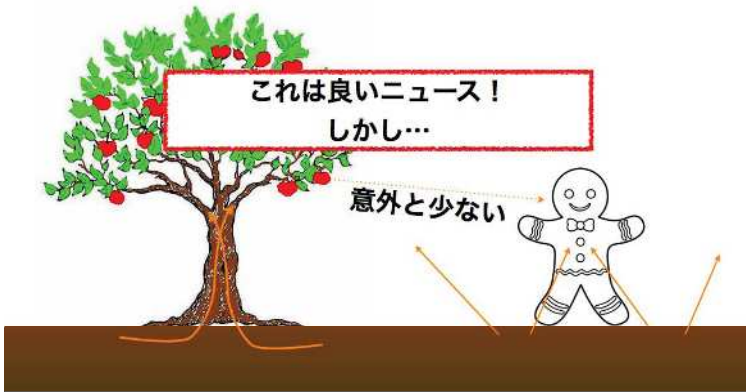
福島原発事故前からイノシシ狩り、山での釣りが趣味

原発事故後も趣味は継続

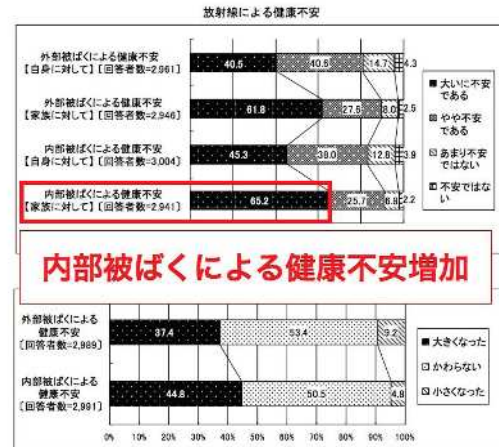
No.	試料名(測定物)	試料量	測定時間(秒)	測定結果 (単位: Bq/kg)			
				減衰補正	①セシウム134	②セシウム137	①+②
1	イワナ	0.0975 kg	3,600	無	50.739	73.519	124.258
					4.5327	5.5563	
2	ヤマメ	0.1015 kg	3,600	無	66.009	93.711	159.720
					5.8107	4.9362	
3	イノシシ肉	0.0975 kg	3,600	無	318.36	515.58	833.94
					8.2442	7.7836	
4	ミョウガ	0.0780 kg	3,600	無	検出されず	検出されず	検出されず
					5.2491	6.6032	
5	ニジマス	0.0935 kg	3,600	無	266.22	425.53	691.75
					7.2880	8.0665	



幸運、規制、測定、生産者の努力



福島市：放射能に関する市民意識調査 (2012.5実施)



食品検査結果にしても
WBC検査結果にしても

全貌を一覧できる
データベースが
存在しない

Part 2

ビッグデータ

Born Digital Materials

東日本大震災
非常に多くの情報が
(デジタル)
で記録された

例 

2011/3/11-17の

日本語のtweet数

179,286,297件 (30GB)

電話が繋がらずSNSが活用される
Twitterは3月11日, 過去最大の177Mツイート/日



一方, こういうデータも沢山
もとはデジタルなのに...

印刷 → 手入力 (時々転記ミス) → PDFで公表

もとはデジタルデータ

1 食品の放射性物質検査について

No.	品名	産地	検査日	検査機関	検査結果	備考
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

厚労省のPDF - 手作業のため誤りが多い

青森県 と殺場採取 (未修正) 横浜検疫所 (修正済)

1 自治体から入手した放射性物質検査結果

No.	検査品名	検査品	検査日	検査機関	検査結果	備考
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

● (平成24年05月22日06時現在)

検体識別番号	出生の年月日	産別の別	母牛の検体識別番号	種別 (品種)
1253160374	H 21.05.24	赤野 (雄)	1241596529	尖蹄種 (肉専用種X乳用種)

飼養場	飼育内容	飼育年月日	飼養場所在地	氏名または名称
1	青森県 出生	H 21.05.24	上北郡六ヶ所村	木村 博
2	青森県 転出	H 21.07.23	上北郡六ヶ所村	木村 博
3	青森県 搬入	H 21.07.23	上北郡七戸町	青森県地域家畜市場
4	青森県 取引	H 21.07.23	上北郡七戸町	青森県地域家畜市場
5	青森県 転入	H 21.07.23		
6	青森県 転出	H 23.08.25		
7	青森県 搬入	H 23.08.25	十和田市	十和田食肉センター
8	青森県 と畜	H 23.08.26	十和田市	十和田食肉センター

http://www.mhlw.go.jp/stf/houdou/2r985200001n9ix-att/2r985200001n9mo.pdf

結果 判明日	結果(Bq/kg)		
	Cs-134	Cs-137	Cs合計
H24.5.9	<0.226	<0.246	<0.47
H24.5.7	<4. 7	<4. 9	<9. 6
H24.5.7	<5. 5	<4. 6	<10
H24.5.11	<1.0	<1.0	<2.0
H24.5.11	<1.0	<1.0	<2.0
H24.5.11	<1.0	<1.0	<2.0
H24.5.11	http://www.mhlw.go.jp/stf/houdou/2r985200002a9wp.html 1 検査結果 417・418		

全角小数点は
半角に

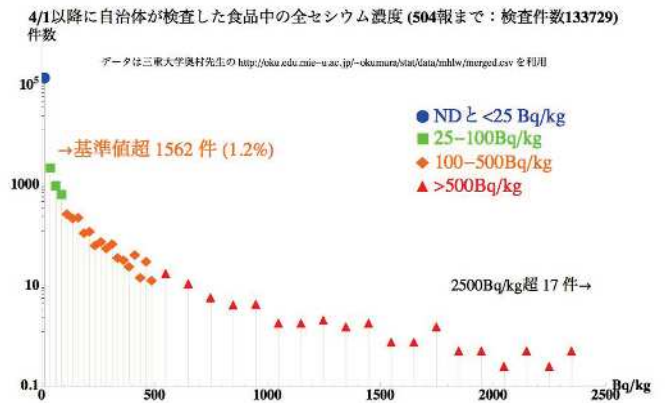
厚労省のPDFをデータ化して検索できるようにした サイトを食品流通構造改善促進機構がボランティア で運営していたが...



ボランティア(三重大奥村先生)が PDF→CSV→検索可能なDBに



それをまたボランティア(早野)が グラフ化してtweet



折角もとはDigitalなのに

そして 失われつつあるデータも

福島原発事故による放射線 Data の Digital Archiving
Digital Archiving of the Data of Radiation
due to the Reactor Accident in Fukushima

政池 明[Ⓐ]
京都大学名誉教授

東日本大震災ビッグデータワークショップ Project 311

- 概要
- 参加申し込み
- 提供データ
- スケジュール
- ディスカッション
- FAQ
- 規約一覧
- プロジェクト一覧
- 開催場所及び費用
- イベント

概要

2011年3月11日の東日本大震災発生時、ソーシャルメディアとマスメディアを通じて大量の情報が広がり、正しい情報もあれば、様々な噂やデマも飛び交い、情報がいかに重要かを改めて知る良い機会でもありました。では、こうした情報はどのように伝えられたのでしょうか？ また、本当に伝えられた情報がなぜ伝えられなかったのでしょうか？ 当時はデータで振り返った時に、本当に必要なサービスは何だったのか、次の災害に備えるために、今我々はどんな準備をすればよいのでしょうか？

『東日本大震災ビッグデータワークショップ - Project 311 -』では震災発生から1週間の間に実際に発生したデータを参加者に提供いたします。参加者はそのデータを改めて分析することによって、今後起こりうる災害に備えて、どのようなことができるかを議論し、サービスを開発することができます。

提供データ

- 3月11日から1週間の朝日新聞記事 (提供: 株式会社朝日新聞社)
- Google Trends (提供: グーグル株式会社)
- 東日本大震災直後のテレビ放送テキスト契約データ (提供: JCC株式会社)
- 3月11日から1週間のツイート (提供: Twitter Japan 株式会社)
- NHK総合テレビ大震災発生直後から24時間の放送音声書き起こし及び抽出ワードランキング (提供: 日本放送協会)
- Honda インターナビ進行実績マップデータ (提供: 本田技研工業株式会社)
- レスキューナウの鉄道運行情報/緊急情報/被害状況のまとめ情報 (提供: 株式会社レスキューナウ)
- 混雑統計データ (提供: 株式会社ゼンリンデータコム)

あと
8
日で
報告会

プロジェクト一覧

Project HAYANO

マスメディア・カバレッジ・マップ

異種協調型災害情報支援システム実現に向けた基盤技術の構築

Twitterデータのクリーニングプロジェクト

位置情報付きのtweetリスト作成プロジェクト

Twitterからの被災時の行動経路の自動抽出およびその信頼性の検証

首都圏における帰宅困難者のモデリング

渋谷プロジェクト

震災データからの情報抽出ツール開発プロジェクト

災害対応質問応答システム&言論マップ

ツイッターデータの意味的解析による災害情報拡散の分析

中間被災地におけるTwitter情報発信の分析プロジェクト

「トレンドキーワード抽出システム」開発プロジェクト

大災害時における特別な支援ニーズを持った被災者に対する情報提供に関するプロジェクト

ハッシュタグの利用傾向調査

流言クラウド: 災害時の医療情報に関する分析

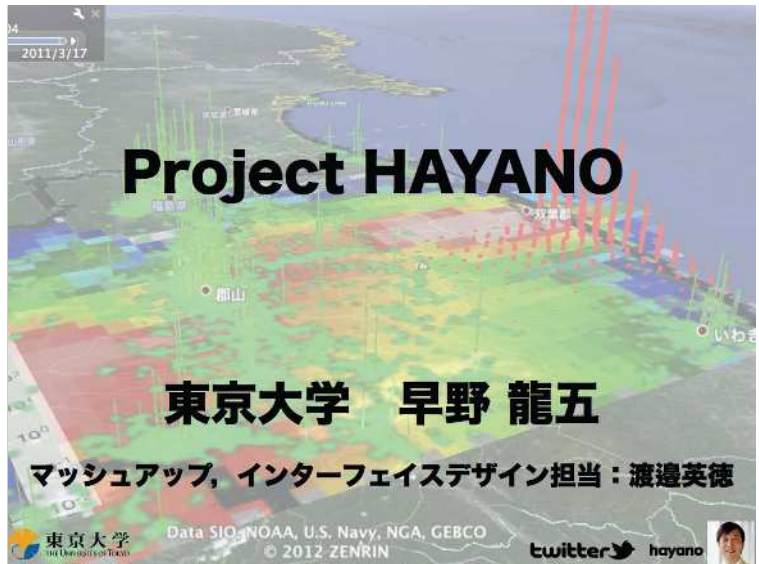
避難タイミング決定行動を調べる Twitter データ編 混雑統計データ編

沿岸被災地復興支援検証・予備調査

Falconseedを使った分析プロジェクト

提供データ

- JCC 株式会社
- Twitter Japan 株式会社
- 株式会社 朝日新聞社
- グーグル株式会社
- 株式会社ゼンリンデータコム
- 本田技研工業株式会社
- 日本放送協会
- 株式会社レスキューナウ



Project HAYANO

目的:

- 放射性ヨウ素による内部被ばくの (オープンな) 評価
- (SPEEDIの有効性の検証 - 未着手)

方法:

各地点でいつ 何人が どれだけ ヨウ素を吸ったか:

SPEEDI (等) × ZENRIN 提供データ

ゼンリンデータコムの 混雑統計データ

オートGPS携帯から5分毎に取得

加速度センサーを用いる (移動がない場合はデータを取らない)

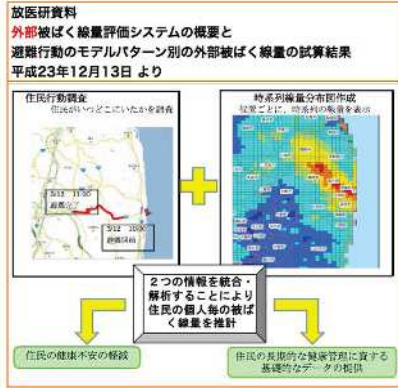
250m×250mメッシュ内の1時間ごとのデータとして提供

背景：

1. 警戒区域等では、事故発生初期に有意に放射性ヨウ素による内部被ばくを受けた可能性

2. 実測データは少なく内部被ばく線量評価への適用は困難

3. 住民行動調査は開示されない(個人情報) → 第三者検証は不可能



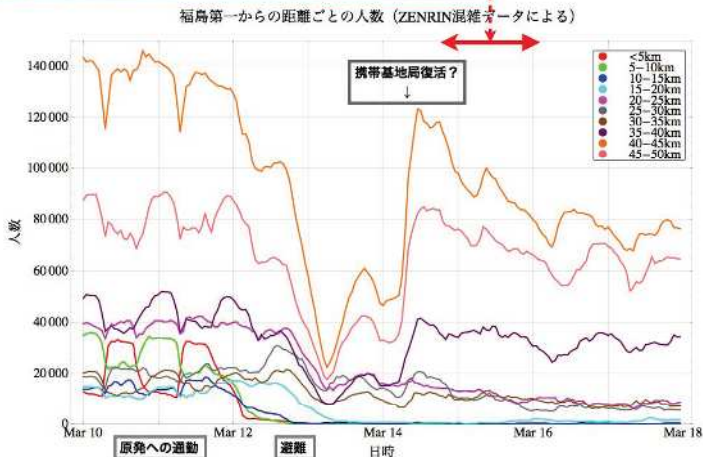
1. ¹³¹I シミュレーション結果を4種入手

- i) SPEEDI - 3月15日以降については、実際のモニタリング測定結果を再現できていない
- ii) 海洋研究開発機構 (JAMSTEC) - 動画
- iii) 国立環境研究所 - 最新の放出源推定を用いた数値データ (これを主として使用)
- iv) JAEA - 動画のみ、数値データの提供には応じていただけなかった

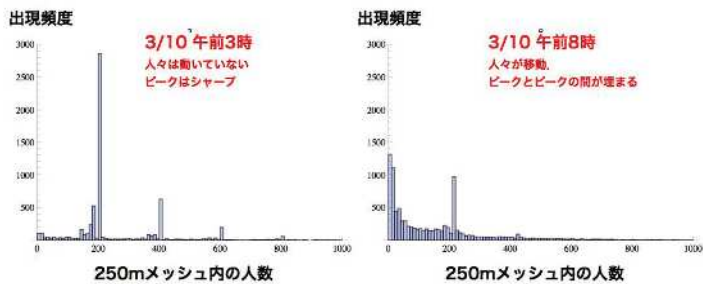
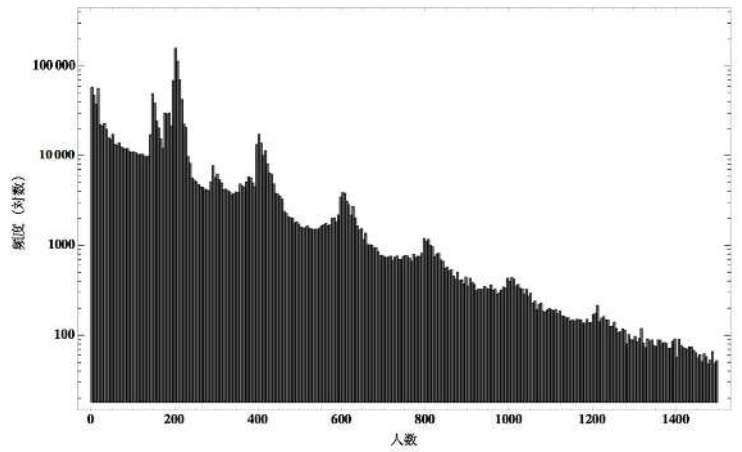
2. ZENRINの混雑データと重ね

3. 地域ごとの 線量 × 人数 を評価中

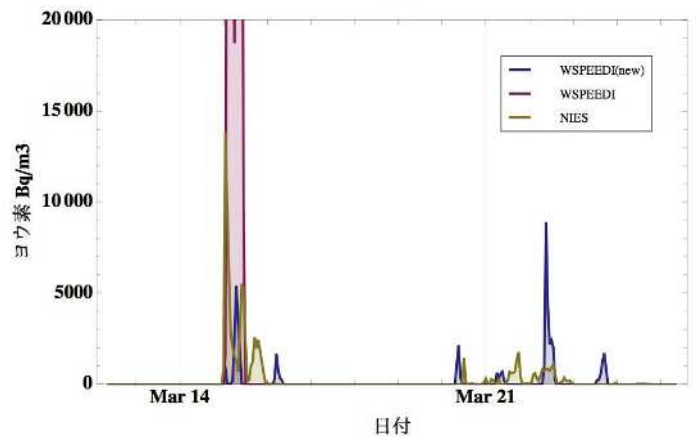
幸いにも ヨウ素濃度が最も高かった時の ZENRINデータは使えそう



ZENRINのデータをじっと眺める

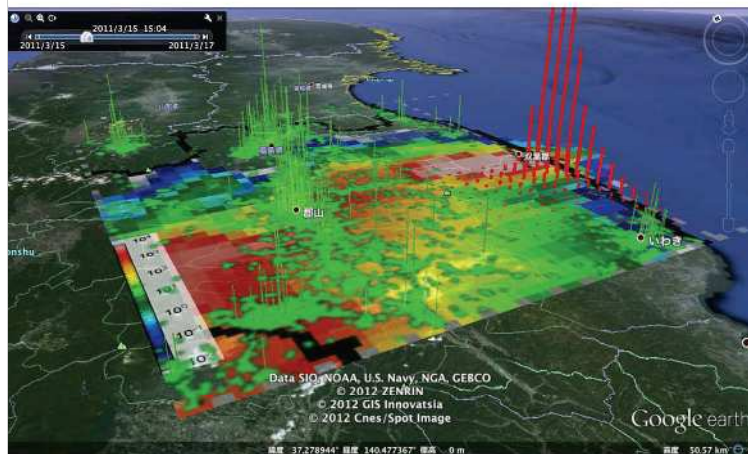
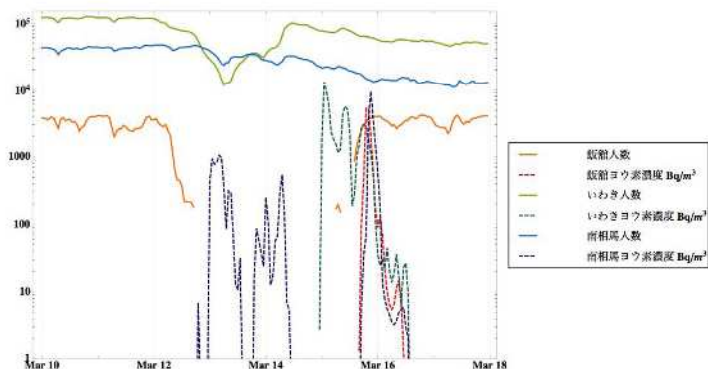


いわき



代表的な例

飯館村, いわき市, 南相馬市の市役所 (村役場) から
半径5km圏内での 人数 vs ヨウ素濃度 (縦軸対数注意!)



10/23 東北市長会決議
(議員立法を目指す)

...

よって、甲状腺検査の対象者である平成23年3月1日時点で18歳以下の県民が18歳を超えて甲状腺がんを発症した場合においても、原子力発電所事故との因果関係を問うことなく、国が治療費を負担するよう強く要請する。

とわらわけて

- ▶ きちんとデータを取り, それに基づいて議論しよう
- ▶ 電源, 通信インフラの重要性
- ▶ もともとデジタルなデータ, デジタルのまま解析・公開・アーカイブ出来る社会インフラが必要
- ▶ 実は (ITと関係ないところで) 人が足りない

ビッグデータは
人を救えるか?