Not What You Think:

A Simple Approach to Scalable Access of CMS Data

National Committee on Vital and Health Statistics

-

CMS Line of Service

Joshua Rosenthal, PhD

PROBLEM

Problem is not security, privacy or man-hours

Problem is a systemic business change that breaks historic infrastructure

-

Yes, it's costing you money to create custom files

But you're moving to a P4P system

You're telling plans, providers, etc. that they need to be data driven

So you need to make data accessible

System-wide CMS data:

Today, research & quality control; STAR & reimbursement data

Tomorrow, Medicaid, etc.

SOLUTION

Counter-intuitively, it's easier to solve all at once

1) Research & Quality Control

Small number of large data sets

[Security, Privacy & Automation]*

2) STAR & Reimbursement

Large number of small files

[Coherency, Usability & Access]**

Solve today's problems with the technical foundation for tomorrow's biz paradigm

* This is pretty obvious;
** This is less obvious but equally important, happy to provide extensive, detailed examples upon request – (Cf. APPENDIX)

APPROACH

Two Paradigms – select or sequence (build & deploy in phases)

Classic: Data Extract System & Tool

- Simple, easy, works, proven & reliable
- Data system centralized & structured; Tool controls access, creates files on demand, etc.
- Solves Security, Privacy & Automation of large Research/Quality data sets
- Solves Coherency, Usability & Access of disparate STAR/Reimbursement files

New-Fangled: Data Explorer

- Cloud-based (note: “cloud” tech not only makes you seem cool but has real benefits)
- Do your work in cloud (upload data to platform vs. pull down – e.g. Google Data Explorer)

EXAMPLE

1) Data Extract System & Tool
   - Access standardized extracts & create custom extracts on the fly
   - GUI-based system (define file: what do you need, what grain, what time)
   - Security through user authentication, permissions
   - Pull/Push (API & RSS concept: Is there new data or not vs. never knowing)
   - Event notification via (RSS/email: "Data has become available")
   - Reduce processing (pull deltas vs. everything repeatedly with adjudication /claims reconciliation internally processed at APCD level [e.g. 3 types of claims from client: New/Update/Cancelled])

2) Quality Control Metrics
   - Tool checks basics vs. current data quality issues (i.e. based columns (e.g. pick columns, add up – check sum)
   - Verification (e.g. universal constants [i.e. shouldn't ever have more than 500 contracts issued])

3) Taxonomies & Hierarchies – (Cf. APPENDIX)
   - Can't be completely flat
   - Need definitions (ID is not a suitable name for column)
   - Standard definitions (vs. 2008 "org name" = contact; 2009 "org name" = something else)
   - Applied Standardization (e.g. one standard way of communicating data problems [missing, bad, too small to report] vs. current [Missing data sometimes blanks, three sentences in data sets])
   - System-based retrospective reconciliation (data extract system makes changes)

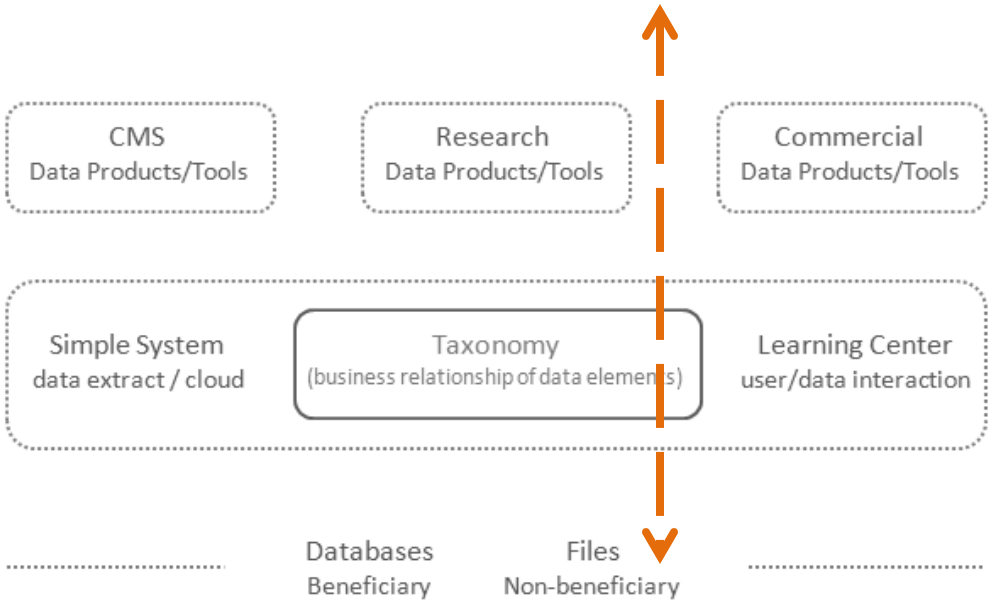
4) Learning Center (Documentation & Community Discussion Boards) – (Cf. APPENDIX)
   - Navigation; Ecosystem tool; Single source vs. current PDFs (makes information inaccessible)
   - Share; Board, FAQ; Searchable deep documentation
   - Best practices; Use cases; "What does this not mean"
   - Integration with data extract tool & notifications

5) Same System for Public Data (STAR, etc.) – (Cf. APPENDIX)
   - Multiple grains (member level expression & non member level [STAR, etc.] – exact same thing)
   - Don't need quick crunching but put the other grains & files in the same ecosystem (tool)
   - Clearly document data sources (understand why X is this number here, vs. y there – at least source list)

INNOVATION

High barrier to entry, tough to share & build insight

Few companies, research groups can do this



(& those who can may have vested interests not to make access easier)

VISION

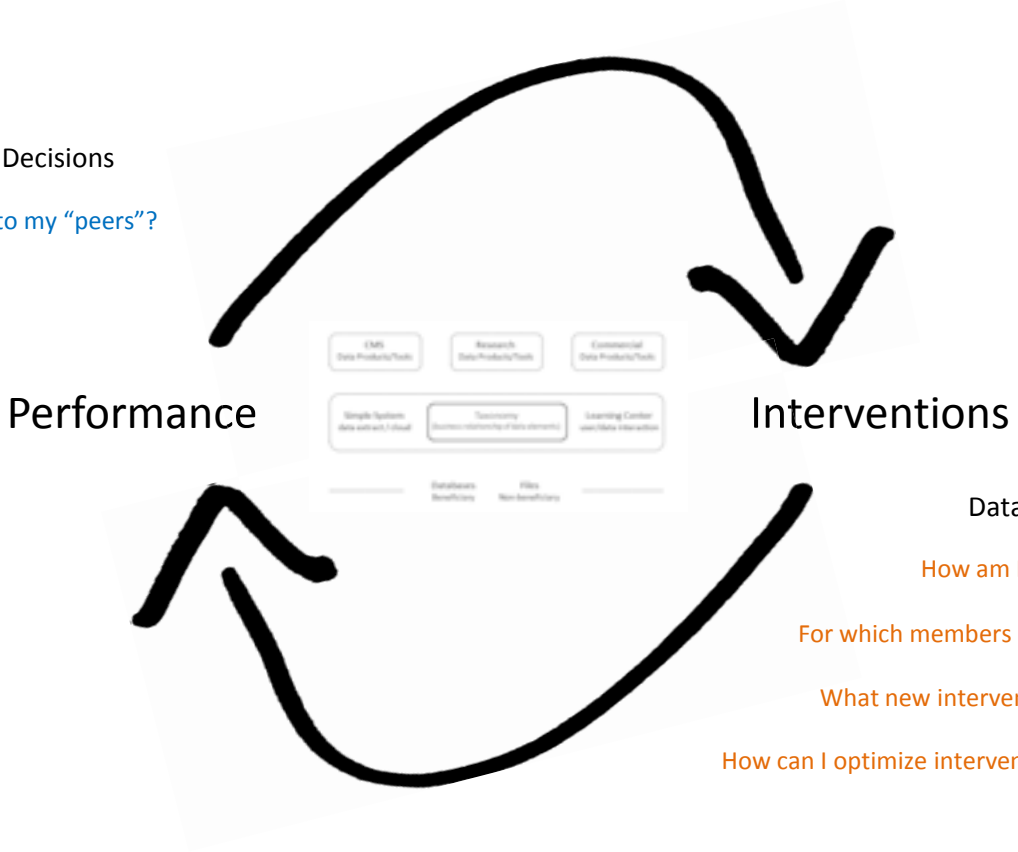
Data-driven Payer (et al) Decisions

How am I doing compared to my “peers”?

Where should I focus?

What should I do?

Which interventions?



Data-driven Vendor Interventions

How am I doing compared to my “peers”?

For which members does my intervention work best?

What new interventions to maximize performance?

How can I optimize interventions to maximize performance?

Data-driven Research

What do successful Payers et al look like ? (geography, supply, socio-demographic, benefit configuration, etc.)

What do successful Interventions look like? (geography, supply, socio-demographic, channel, etc.)

What are the characteristics and key drivers of improvement?

How can policy stimulate these key drivers?

PERSPECTIVE

Joshua Rosenthal, PhD

Melanie Rosenthal

Burak Sezen

Entrepreneurs – Multiple successful exits

Data & Analytics – Designed, built, sold & scaled systems and tools

Awards – DMAA, Business Week, Entrepreneur Mag., etc.

Recently – Founded startup; deep in CMS data

Yale; Sorbonne (Applied Institute for Advanced Studies); Fulbright;
Solstice Capital; Pricewaterhouse Coopers; Health Dialog (first employees);
Continuity of Care Record (CCR) Data Standards Committee; Human Genome Project; Google Health Partner;
Disease Management Association of America (DMAA) Best New Product Winner; Entrepreneur Magazine, 100 Companies to Watch:
Speaker @ SXSW Interactive, World Health Congress, Tableau Data Visualization & Business Intelligence Conference,
Health Care Executive Leadership Network, Health 2.0, Department of Health and Human Services-Institute of Medicine &
National Academies of Science Health Data Initiative, H@cking Medicine, MIT Entrepreneurship Center;
Guest Lecturer @ Harvard College Society of Biological Engineers, The Harvard College Entrepreneurship Forum,
The Harvard Biotechnology Association, The Harvard Healthcare Innovation Club;
Expert Advisor on Health Care Technology Innovation to the Office of the National Coordinator for Health Information Technology

APPENDIX:

Taxonomy

-

How to make data meaningful
for business problems and performance questions

- and -

How to spur innovation
from both start-ups and status-quos

KEY

Before data products, tools, widgets, benchmarks, APIs, systems…

Before access and security…

Before automation and efficiency…

Taxonomy

… basis for moving to new paradigm from fee-for-service to performance based payment

… bridges business & technology/data… bridges internal & external users

... across CMS… identifiable beneficiary databases and non-beneficiary (STAR, et al) files

Internal for CMS use, then exposing outward

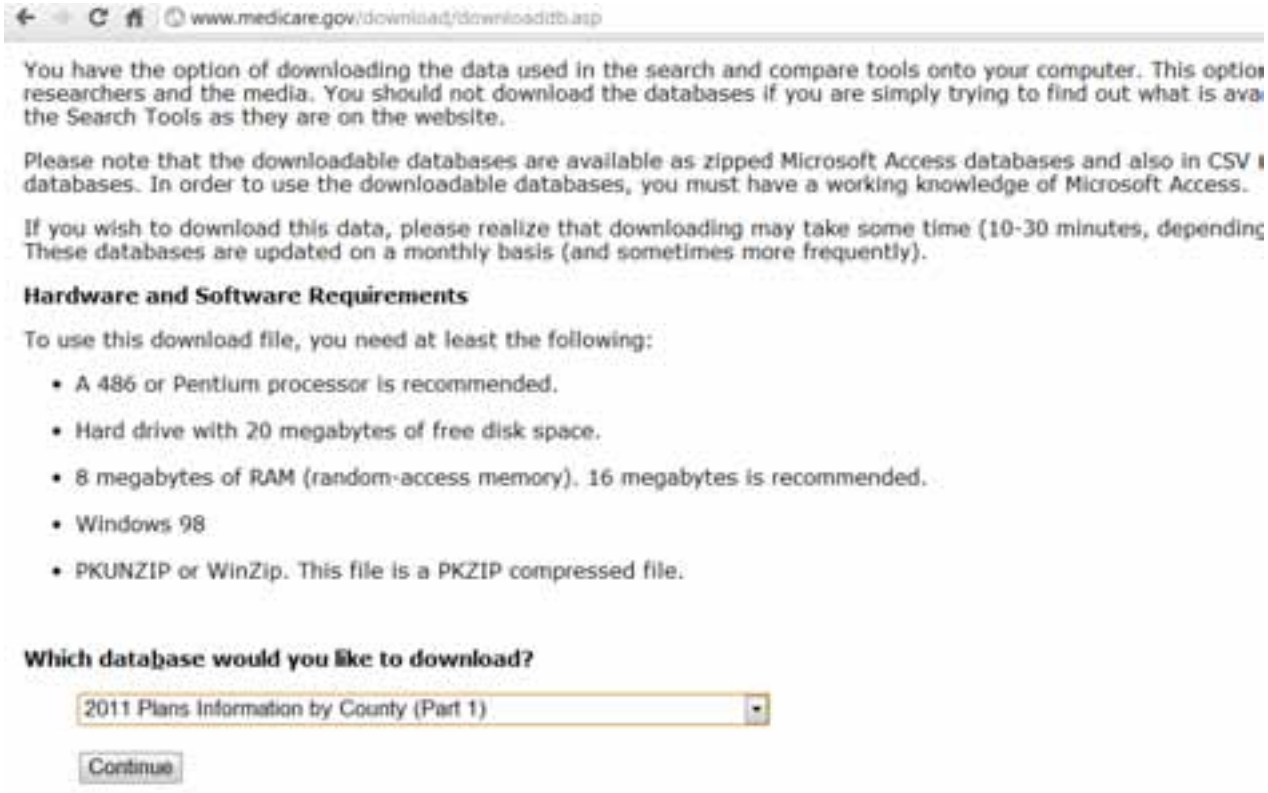
First move, tactically, technically, strategically

Not difficult; achievable (really, it is - flip forward)

SPECIFICS

Example taxonomy (non-beneficiary / STAR example)

http://www.medicare.gov/download/downloaddb.asp



SPECIFICS

What is this? How can I use it to answer business/performance questions?

ContractNum	PlanID	SegmentID	ContractYear	OrgName	PlanName	SPPlanName	GeoName	TaxStatusCd	TaxStatusDes
"E5088"	801	0	2011	Deseret Mutual	Deseret Secure (Employer PFFS)		National Area	2	Non-Profit

SPTaxStatusDesc	PlanType	PlanTypeDesc	CvrgGapOfrd	CvrgGapInd	CvrgGapDesc	ContractImportantNote	SPContrac
Sin Ganancias	40	Employer/Union Only Direct Contract PFFS	FALSE	0	No Gap Coverage		

WebAddress	PartDWbAdrL	FrmlryWbstAdr	PhrmcyWbstAdr	FedApprovalStatus	SPFedApprovalStatus	POSAvailFlag	MailOrderAvail
www.dmba.com				Approved by Medicare	Aprobado por Medicare	FALSE	FALSE

PlanImportantNote	SPPlanImportantNote	SegmentImportantNote	SPSegmentImportantNote	LegalEntityName	TradeName
				DESERET HEALTHCARE EMPLOYEE BENEFITS TRUST	

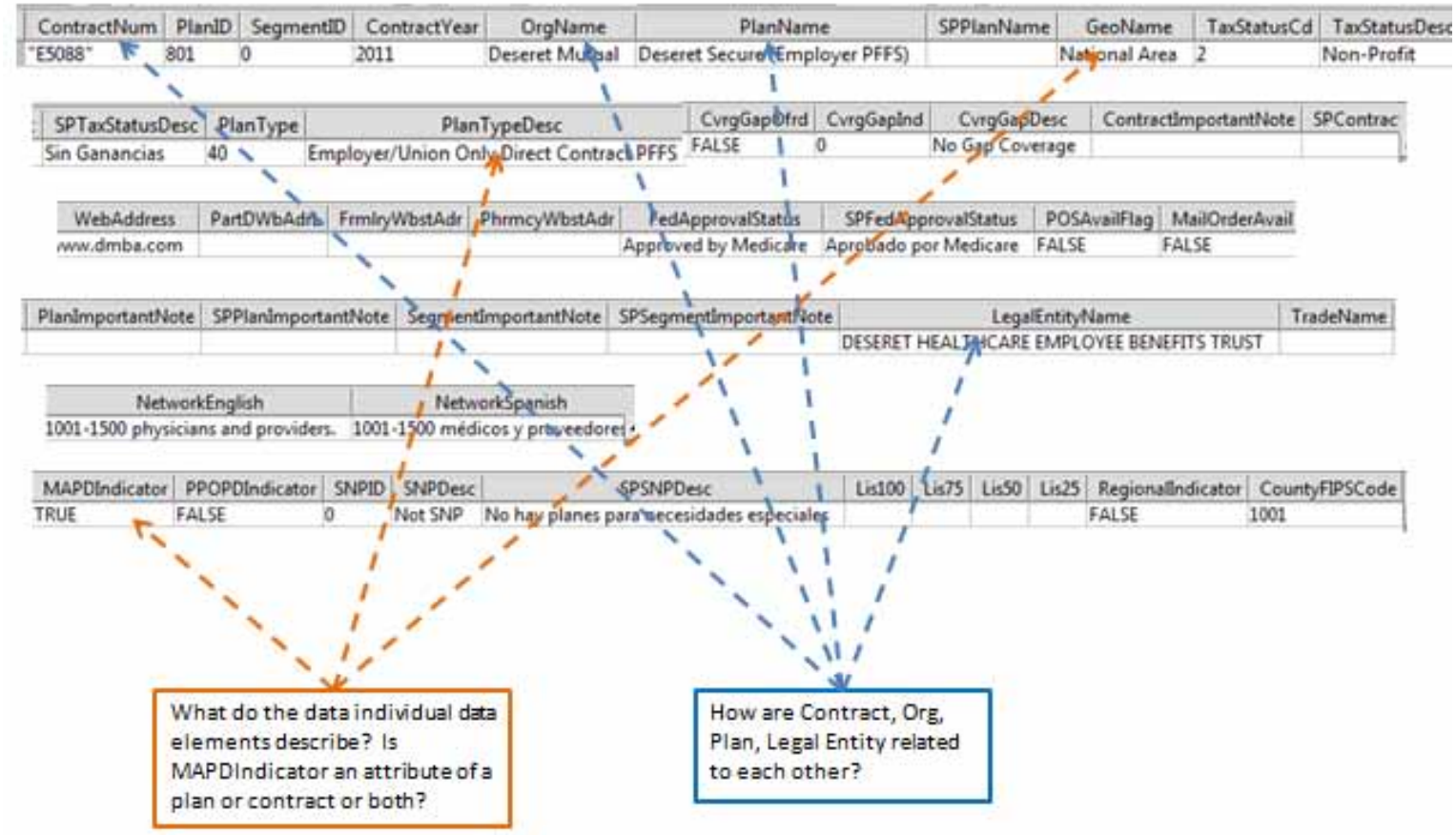
NetworkEnglish	NetworkSpanish
1001-1500 physicians and providers.	1001-1500 médicos y proveedores

MAPDIndicator	PPOPDIndicator	SNPID	SNPDesc	SPSNPDesc	Lis100	Lis75	Lis50	Lis25	RegionalIndicator	CountyFIPSCode
TRUE	FALSE	0	Not SNP	No hay planes para necesidades especiales					FALSE	1001

Sample record from the downloaded file

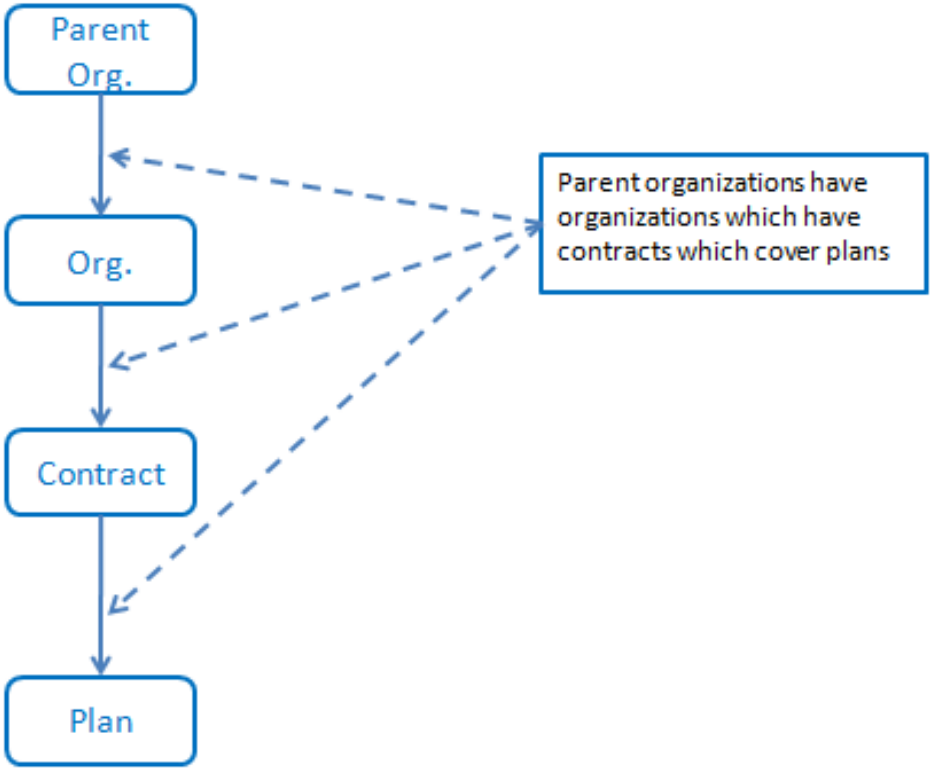
SPECIFICS

In order to get insight, I need the data in a meaningful business structure



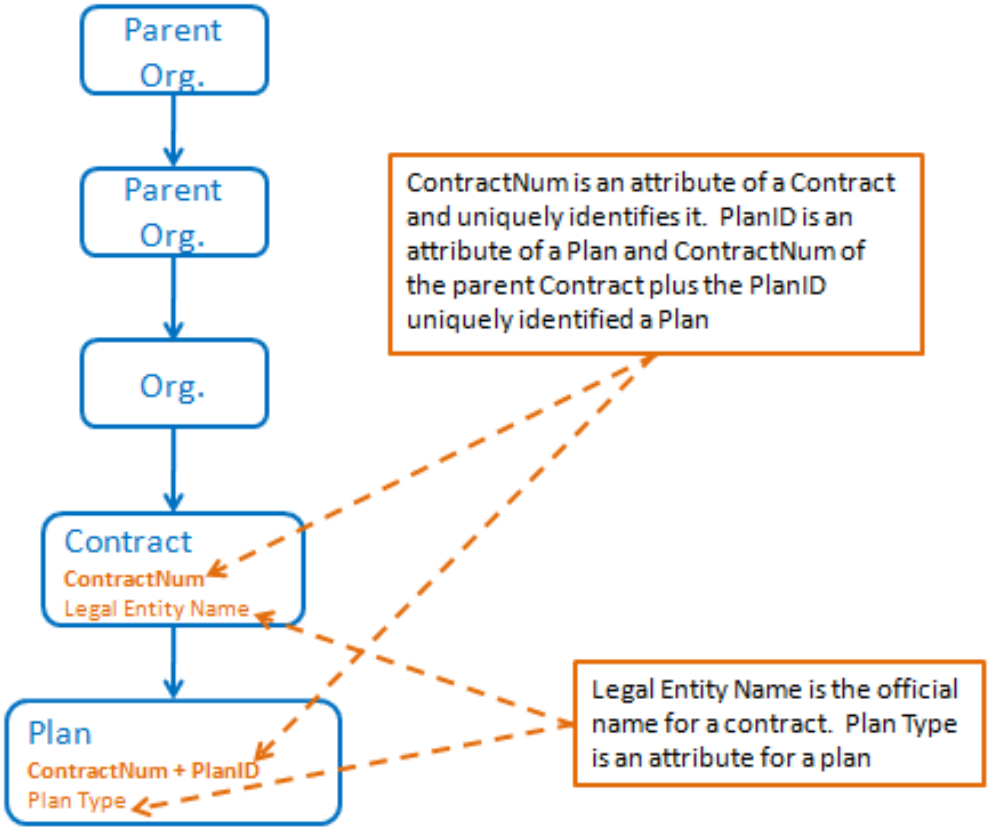
SPECIFICS

Taxonomy defines business entities and the relationships among them

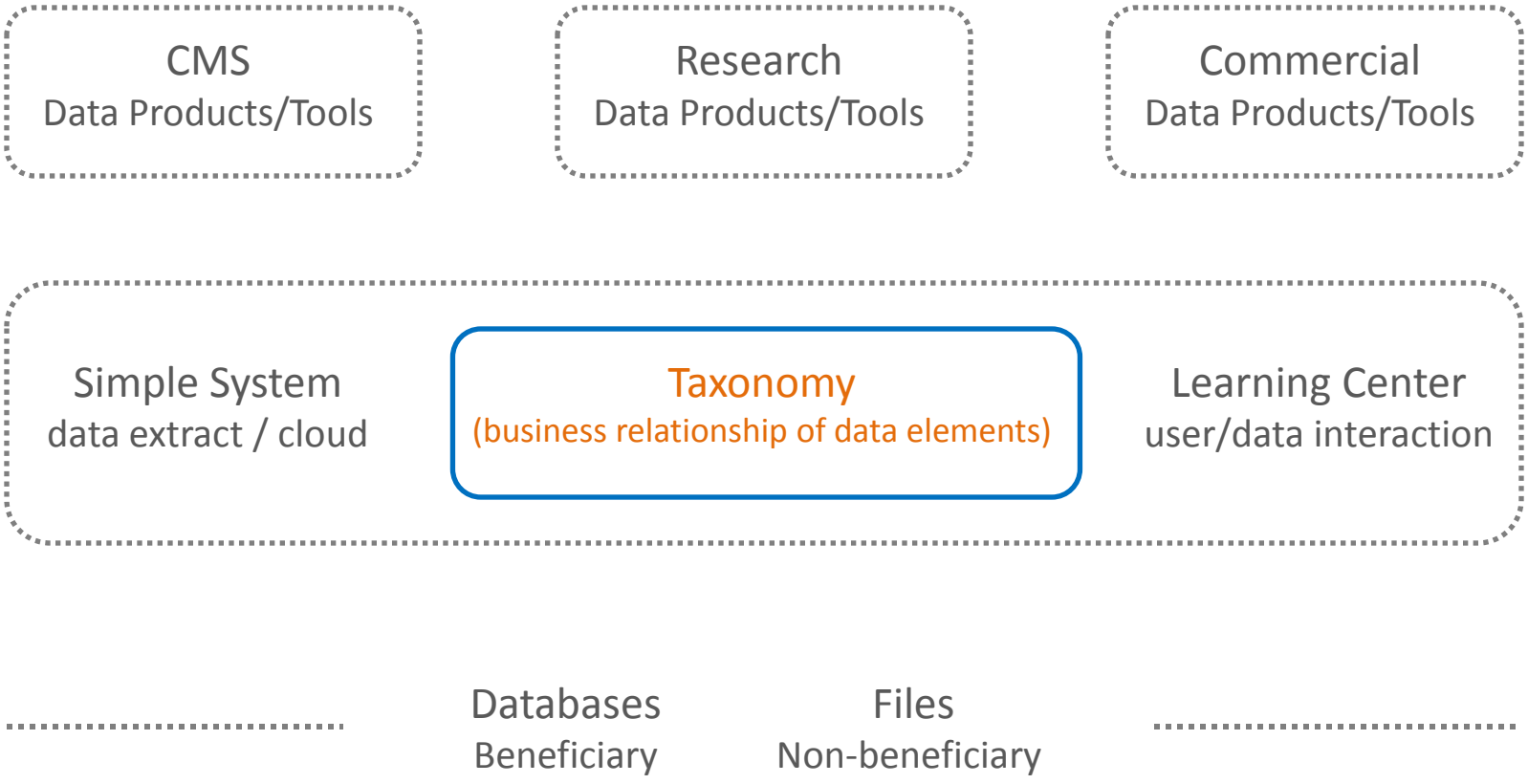


SPECIFICS

Taxonomy defines attributes for business entities

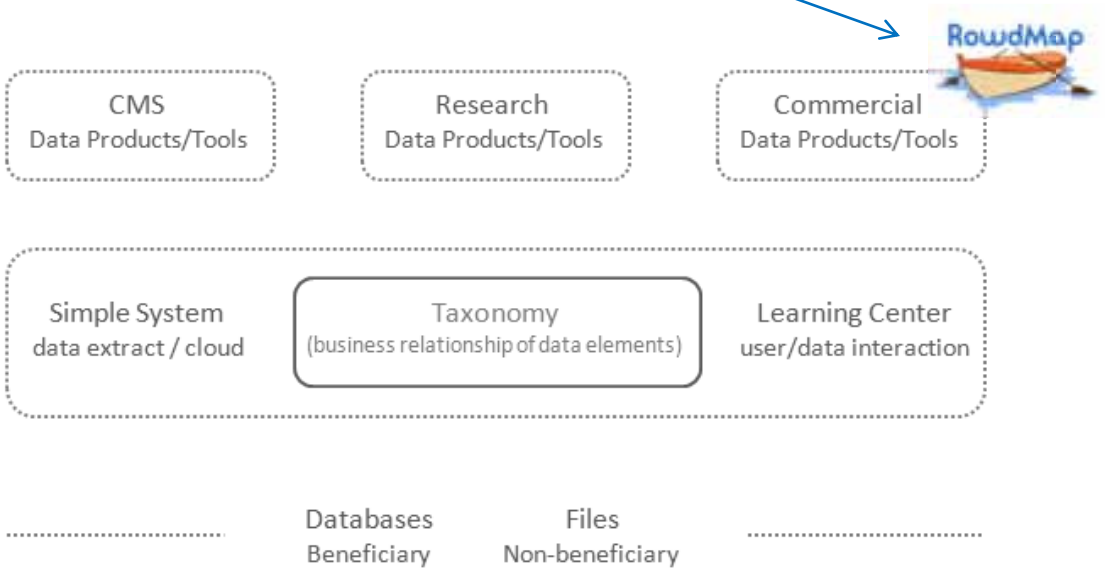


PICTURE


CMS
Data Products/Tools
Research
Data Products/Tools
Commercial
Data Products/Tools
Simple System
data extract / cloud
Taxonomy
(business relationship of data elements)
Learning Center
user/data interaction
Databases
Beneficiary
Files
Non-beneficiary

STORY

Start up (w/deep industry experience/awards/credentials)
wanted to build something cool* here



* Heal Profit Intelligence platform
to measure, project & recommend
strategies and specific interventions
to maximize performance/reimbursement/ROI

STORY

So we had to build it all*

But creating even portions will spur others' innovation & increase performance

How am I doing compared to my "peers"?

Where should I focus?

What should I do?

Which interventions?

CMS
Data Products/Tools

Research
Data Products/Tools

Commercial
Data Products/Tools

Simple System
data extract / cloud

Taxonomy
(business relationship of data elements)

Learning Center
user/data interaction

Databases
Beneficiary

Files
Non-beneficiary

*Non-beneficiary CMS data – taxonomy extends to interventions to measure, predict recommend what actually works
(taxonomy covers additional geography, supply, socio-economic data)

APPENDIX:

Public Learning Center

-

How to get lots of people to use CMS/HHS data
for meaningful things

- and -

How to develop data products
that people want and will use

THESIS

CMS Is Doing Great

Lots of good thing in the works – bigger data Valhalla and interim pub files, challenges etc.

Has done all the hard work and now needs to pivot just a bit to capitalize on massive value

Inundated with various requests need a bit of a filter to screen out noise putting onus on those requesting will help them create value themselves (vs. build dependency on CMS)

No need for data Valhalla – maybe it comes maybe it doesn't, but you are at a tipping point and just need to spin a bit to create real value and capture the return on your investment – and if you get to data Valhalla this will only help you anyway

ISSUES

A Few Things

A) CMS data is difficult for folks to access and understand, especially large audiences of smart folks outside health care

B) Current public data approaches (i.e. public files) are basically academic in nature, divorced from business applicability (hence few users difficult to demonstrate impact)

C) Current innovation approaches (Challenges, etc.) i.e. lots of health start up noise making requests for CMS but very few commercially viable to create self-sustaining public/social/performance change

D) Not enough widespread users from outside health care – new innovative thinking and a broader national discourse (better ROI on your investment – best way for the broader market to 'kick in')

STEP 1

Create a Learning Center

A – ‘Consumer friendly’ web place for people to learn whether experts in CMS data, new entrepreneurs of folks new to the space

B – Can have forms where folks ask and answer questions, deposit and share code, simply share resources – can be open or password, credentialed

C – This is expected in most tech communities and doesn’t’ have to be tough (make it a challenge to build the best one)

D – Start by putting in existing documentation & resources (HDI videos, slides & curriculum, etc.)

CMS / HHS PUBLIC

STEP 2

Take Current Data & Challenges and Tie Them to Real Biz Problems

A – Every challenge etc has to submit the basic business applicability and use CMS / HHS data (either as part of product or outcomes measurement as part of the application & judgment criteria)

B – Curriculum (slides / videos / interviews) available in learning center – about how to do it and from people who have done it (i.e. how to use and monetize data – and even various peculiarities of health care market for folks new to it)

C – Samples of challenge /data / biz applicability available in the learning center

D – Additional data sets (census, gov.org) available in the learning center

BIZ APPLICABILITY OF CHALLNGES & DATA

STEP 3

Plug in Current & Planned Public Data (Files, Models, Tools)

A – Public data file per NORC/IMPAQ – public / synthetic file (current) – even a slightly retooled one geared toward biz problems (i.e. keep county level so this can be tied to STAR contracts for payer performance even if it means getting rid of something else to meet re-identification criteria while retaining specific biz applicability)

B – Artificial file (for companies to use to configure their systems – for any public user to play with to learn the data structures)

C – Any basic data model (not the process of de-identification but the basic taxonomy, entities relationships, etc – and entity diagram or even conceptual model

D – Any basic meta data or even meta data tools (exploration, parsing etc.)

FILES & DATA MODELS / TOOLS

STEP 4

Take an Applicable File and Put the Data in a Relational Model, Then Push to a Public Data (Analysis) Explorer

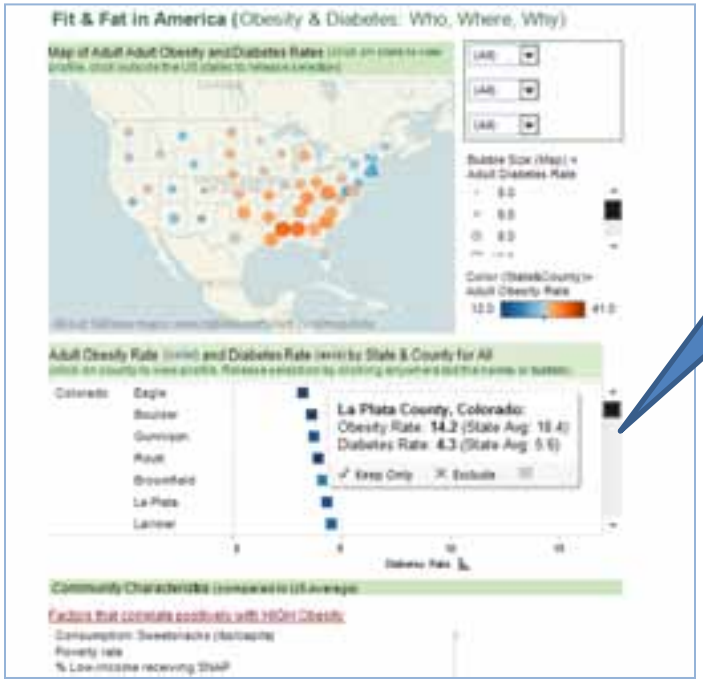
A – Most likely an artificial file (kind used by companies to prime systems before receiving CMS data) but could be a public synthetic (either credentialed access or a special versions further de-identified)

B – Build a basic taxonomy (relationship of data entities to each other an biz questions) – this could be done as a challenge

C – Push to various public data explorers (Google data explorer, tableau public, etc.)

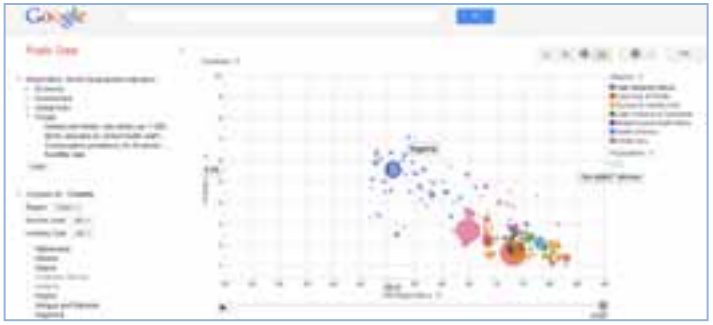
D – Potentially partner & publicize

EXPOSE DATA IN CURRENT EXPLORERS



Taxonomy controls relationships and allows users to analyze

CMS data pushed to current explorers, Users interactively analyze and share



STEP 5

Partner for Users and Monitor the Usage

A – Partner for challenge with broader tech communities (cf. Google, O'Reilly, et al - NB the challenge that won the Read Write Web / Tableau challenge was on co-morbidities for diabetes using government data: http://goo.gl/Akb7M)

B – Monitor usage (links in and out and specific file analysis and usage via basic analytics – use it for feedback in addition to surveys/round-tables/calls – as well as learning center reading/comments/view)

C – Development CMS data products based on usage (both from the business applicability of data & challenges as well as the curriculum/learning and the specific file use and file exploration in the data explorer environment)

PARTNER & DEVELOP DATA-DRIVEN DATA PRODUCTS

Partner & Publicize

Monitor site, resource, data & tool usage

Allocate resources and develop data products based on value & usage

Hmm, most popular data / content is the Todd Park photo gallery. So build a product :-)