PROJECT

DataSphere

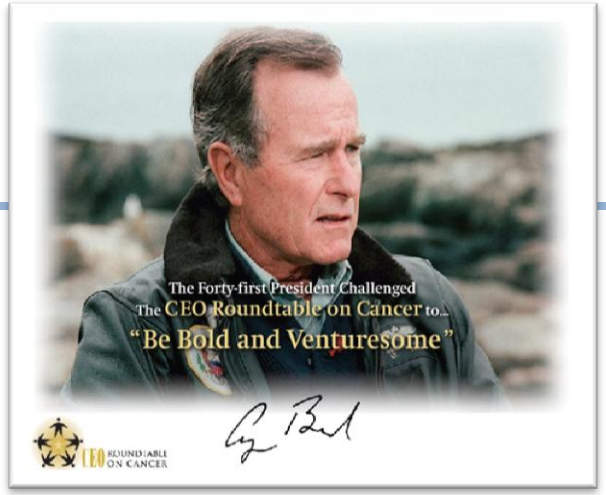
An Initiative to Make Academic and Commercial Data Sharing Happen in Cancer Research

Charles Hugh-Jones MD MRCP, Sanofi Oncology
on behalf of CEO Roundtable on Cancer, Life Sciences Consortium
Institute of Medicine
October 4th 2012

an LSC initiative

Disclaimer: Views expressed are personal and not necessarily those of Sanofi Oncology

CEO Roundtable on Cancer



- "Life Sciences Consortium" working team
- Address issues in cancer research
- Accomplish together what no single company might consider alone

COVANCE
THE DEVELOPMENT SERVICES COMPANY
DukeMedicine
gsk GlaxoSmithKline
Johnson & Johnson
THE UNIVERSITY OF TEXAS MD ANDERSON CANCER CENTER
Making Cancer History®
Takeda MILLENNIUM
THE TAKEDA ONCOLOGY COMPANY
MOFFITT CANCER CENTER
NOVARTIS
PPD
QUINTILES
SANOFI
sas
astellas
AstraZeneca
Pfizer
Lilly

Engages 3rd parties as "Safe Harbors"

NATIONAL CANCER INSTITUTE
INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES
FDA

an LSC initiative

www.ceo-lsc.org

PROJECT DataSphere

Agenda: A model for data sharing

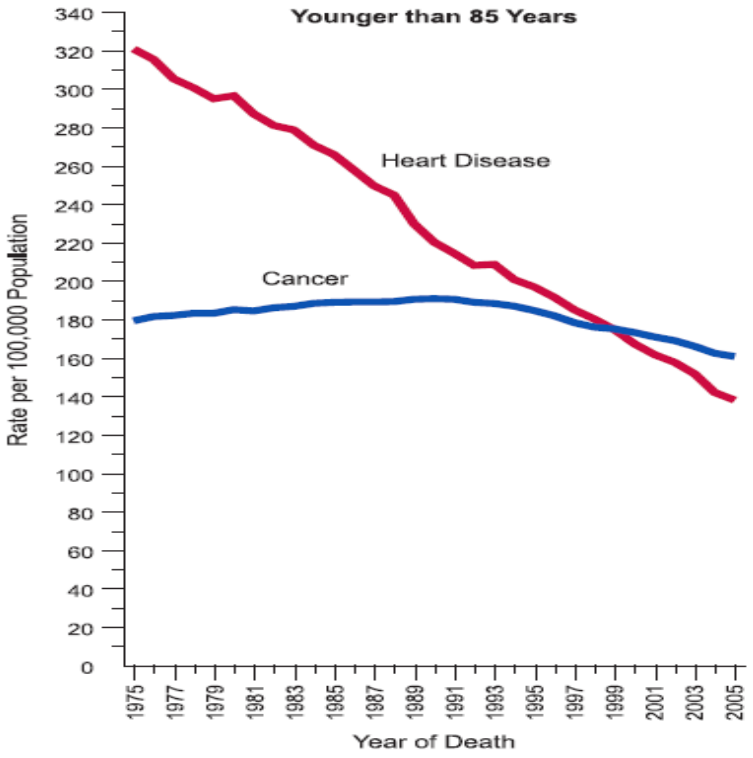
- The Problem with Oncology
- Data Sharing
- Project DataSphere

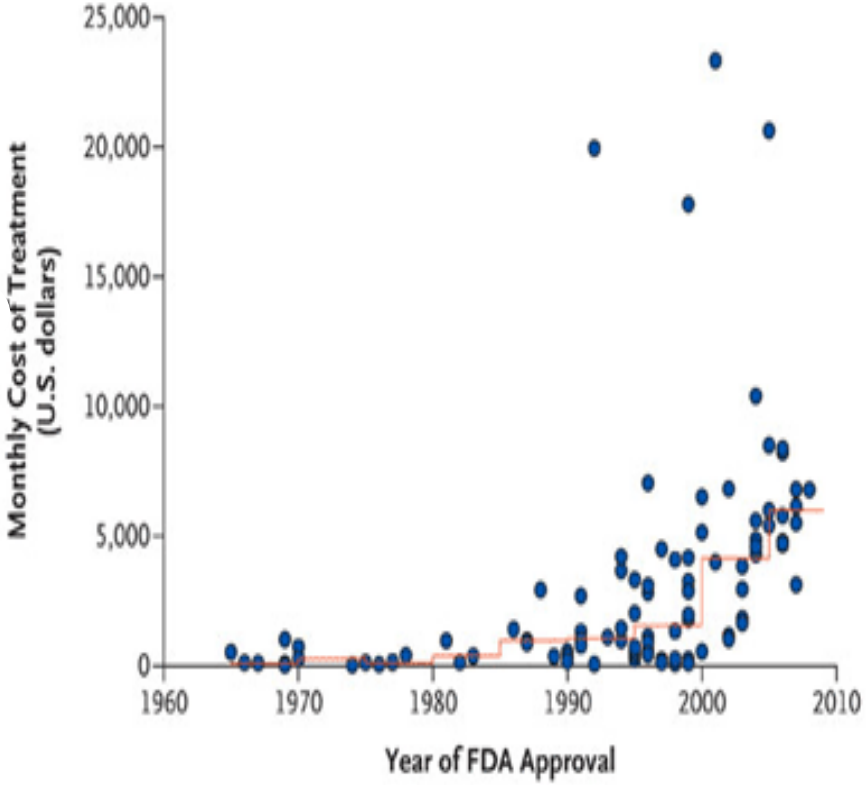
- 9 minutes…

an LSC initiative

PROJECT DataSphere

Cancer Mortality and Cost…





1 Jemal A: CA J Clin 2009

2 Bach, NEJM 2008

an LSC initiative

PROJECT DataSphere

Agenda: A model for data sharing

- The Problem with Oncology
- Data Sharing
- Project DataSphere

an LSC initiative

PROJECT DataSphere

Data Sharing in Oncology : why do it…?[1]

7.6 M lives lost each year worldwide

1. Faster, more efficient research
   - Improved trial design and statistical methodology
   - Secondary hypotheses & epidemiology
   - Collaborative model development
   - Smaller trials sizing (esp. with molecular subtyping)
1. Reproducibility and reduced duplication
2. Transparency, and prevention of selective reporting
3. Real World Data corroboration with Trial Data
4. Unknowns [2]
5. Data Standards & Meta-analysis

an LSC initiative

1: Vickers 2006
2: www.cardia.dopm.uab.edu: 475 publications from a single large dataset

PROJECT DataSphere

Everyone's onboard, but…

VOLUME 29 · NUMBER 9 · MARCH 20 2011

JOURNAL OF CLINICAL ONCOLOGY | E D I T O R I A L S

Clinical Research: Show Us the Data

Alberto Ocana, *Princess Margaret Hospital, Toronto, Canada; Albacete University Hospital and Asociacion Española Contra El Cancer Unit, Albacete, Spain*
Eitan Amir, *Princess Margaret Hospital, Toronto, Canada*
Bostjan Seruga, *Institute of Oncology Ljubljana, Ljubljana, Slovenia*

See accompanying editorial on pages 1091 and article on page 1204

CNBC.com

FDA sees huge opportunities in opening up drug data

Reuters
05 Dec 2011 | 10:53 AM ET

LONDON - Regulators and drugmakers need to find ways to make more clinical data openly available, since vital knowledge about fighting disease is often locked away in confidential databases, the head of the U.S. drugs watchdog said on Monday.

Food and Drug Administration (FDA) Commissioner Margaret Hamburg said opening up data to public scrutiny needed to be done selectively, given legitimate concerns among companies over commercial confidentiality, but more could still be done.

- Vast quantities of sub-optimally used clinical data [5]

  …yet neither industry nor academia are sharing [3]

1: Ocana et al, JCO 2011
2: Peggy Hamburg, FDA Dec 2011 – CNBC webiste
3: Savage & Vickers, 2009. PLoS One
4: PharmaExec 2012: 900 drugs / 8000 trials

an LSC initiative

PROJECT DataSphere

Agenda

- The Problem with Oncology
- Data Sharing
- Project DataSphere

an LSC initiative

PROJECT DataSphere

A Need Exists…so why won't they?

- Unique challenges in healthcare
  - Multiple valid attempts : CaBIG, CTCAP, Biogrid Aus.,Transcelerate
  - Attitude is "don't share unless I can prove no harm occurs"[4]
- Academic Disincentives
  - Academic tenure system driven by data hoarding[1] [2]
- Patient
  - Privacy, Confidentiality, Consent & Ethics [3]
- Corporate
  - IP & Competition Law
- Everyone
  - Resources for data preparation
  - Suitable IT environment

an LSC initiative

1: Kaye *et al* 2009
2: Tucker 2009
3: Westin, IOM 2007
4: Vickers 2006

PROJECT DataSphere

What would “great” be…?

- Simple.
- Systematic sharing
  - Access is initial challenge
- Comparator arms[2], protocols, CRFs & descriptors
  - Industry & academia, positive & negative data
- “Publically” accessible, simple web-library, respecting privacy issues [3]
  - Using convening power of CEOs, Patient Advocacy groups and social media
  - Standardized de-identification, secure IT, DUAs and user tools
- Third party data “aggregation” partners due to breadth of oncology data
  - Too much potential data requires a simple approach for Oncology
- The DataSphere[1]

an LSC initiative

1: Public access projected as April 2013
2: Post concept goal to include active arm and genetic data
3: Access criteria include recognized research institution, data use agreement, and use consistent with data sharing goals

PROJECT DataSphere

Make it easy to do…

1 Why?

- **Incentives**
- **Donor**
  - Productivity[1]
  - Cost savings[2]
  - Citation[3]
  - Collaboration
  - CSR / Micro-A
- **Patients**
  - Using data
  - Secure
- **Researchers**
  - Competitions

2 How?

- **Tools**
- **Donors**
  - Standard de-identification
  - One DUA
  - Resources
  - How to guides
  - "Service"
  - Forums
- **Advocacy**
  - Tools
  - Social Media
  - Rating

3 Where?

- **Environment**
- Secure
- Simple
- Powerful
- Scalable
- Accessible
- Neutral
- Automated
- Tiered (?)
  - Trust differential[4]
- Metrics
- Collaboration

an LSC initiative

1: Paul et al, Nature Rev Drug Disc, March 2010
2: $251MM savings across broad program Internal data on file
3: Piwowar et al PLoS One March 2007

PROJECT DataSphere

- **The primer:**
  - 2 Sanofi-donated recent Phase III datasets/ CRFs - Breast and Prostate cancer
- **The Goal**
  - 30+ high-quality datasets, by priority disease area, by key LSC members end 2013

an LSC initiative

PROJECT DataSphere

DataSphere: Modified Safe Harbor de-identification

Table 2. Project DataSphere de-identification model.	
(A) Names	
(B) All geographic subdivisions are reported at a granularity no smaller than a country	
(C) All ages below 85 will be fixed at the point of trial registration, rounded down to the nearest integer; all ages over 85 reported as 85+	
(D) All death-related events, as well as direct indicators of such an events, are reported at a granularity no smaller than one week	
(E) Telephone numbers	(N) device identifiers and serial numbers
(F) Fax numbers	(O) Web Universal Resource Locators (URLs)
(G) Email addresses	(P) Internet Protocol (IP) addresses
(H) Social security numbers	(Q) Biometric identifiers, including finger and voice prints
(I) Medical record numbers	
(J) Health plan beneficiary numbers	(R) Full-face photographs and any comparable images
(K) Account numbers	(S) Any other unique, identifying number, characteristic, or code
(L) Certificate / license numbers	
(M) vehicle identifiers and serial numbers, including license plate numbers	

- Oncology-needs focused
  - Minimize resources
  - Tools
  - Unified approach
  - Statistician method
    - **More restrictive aspects**
    - No 3 code ZIP & 85+ grouping
    - **Less restrictive aspects**
    - Death to week
    - Visit dates exl. certain SAEs
    - Birth +/- 365 days
- DataSphere methodology
  - 0.00029% of U.S. population unique
  - ~100 x safer that Safe Harbor[1]
  - Consistent with HIPAA, EU etc

1: Malin B, PhD. DataSphere data on file

an LSC initiative

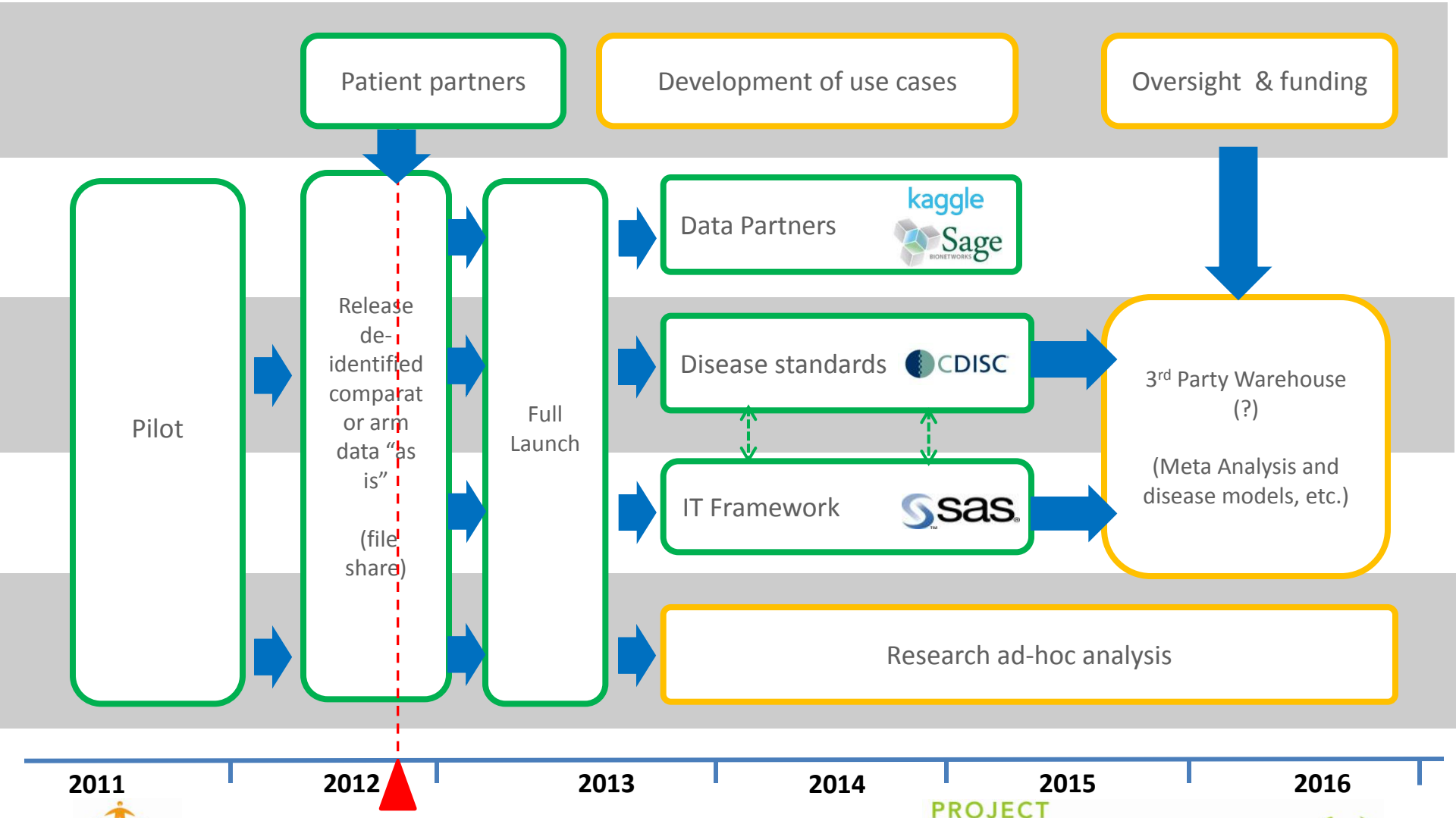
PROJECT DataSphere

Donors & Patients: Change the social paradigm

an LSC initiative

PROJECT DataSphere

What next…?

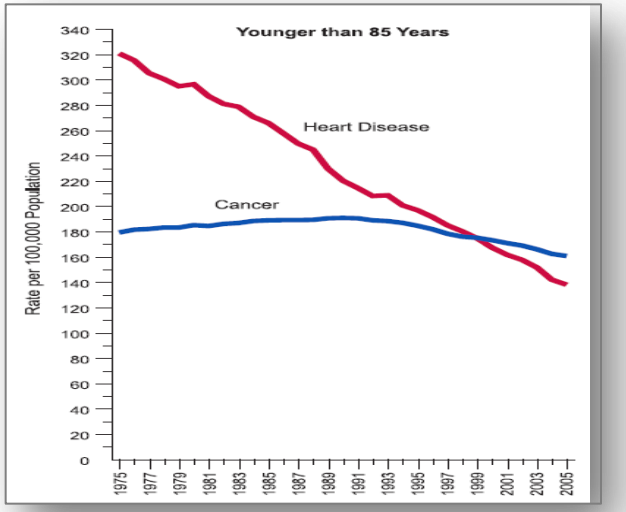


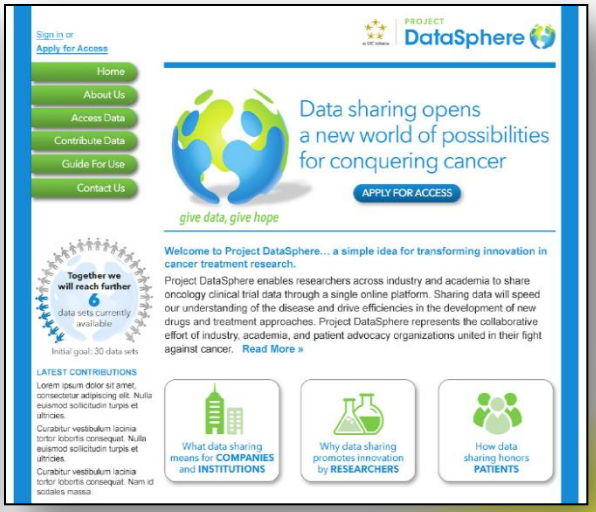
an LSC initiative

PROJECT DataSphere

Critique: What could be better…?

- Concept will define success
  - No active arm nor genomic data facility yet – unique challenges
  - De-identification can never be complete, nor data full
  - Resource challenges and nascent business model
  - Accurately defining ongoing social-media and advocacy-driven components
  - Defining micro-attribution component
  - Long term home for environment
- **KPIs:**
  - Quantity and Quality of Datasets donated
  - Dataset Specific Use Cases
  - Security





an LSC initiative

For information/newsletter: Projectdatasphere@ceoroundtableoncancer.org

Public Access April 2013: Projectdatasphere.com

Acknowledgement

- Project Office: Robin Jenkins, Michael Curnyn, John Dornan
- Legal: John O'Reilly, Anne Vickery, Phil Porter
- Biostatistics: Zhenming Shun, Jeff Cortez, Bradley Malin
- Clinical: Leonardo Nicacio, Ronit Simantov, Stephen Friend, Amy Abernethy
- IT: Jeff Cullerton, Mark Kwiatek, Angela Lightfoot, Janice Neyens, John Brocklebank
- Advocacy: Joel Beetsch, Deb Sittig, James Shubinski, Nicole Johnson, Cindy Geoghegan
- Sponsors: CEO Roundtable on Cancer leadership
- And many others….

PROJECT

DataSphere

an LSC initiative