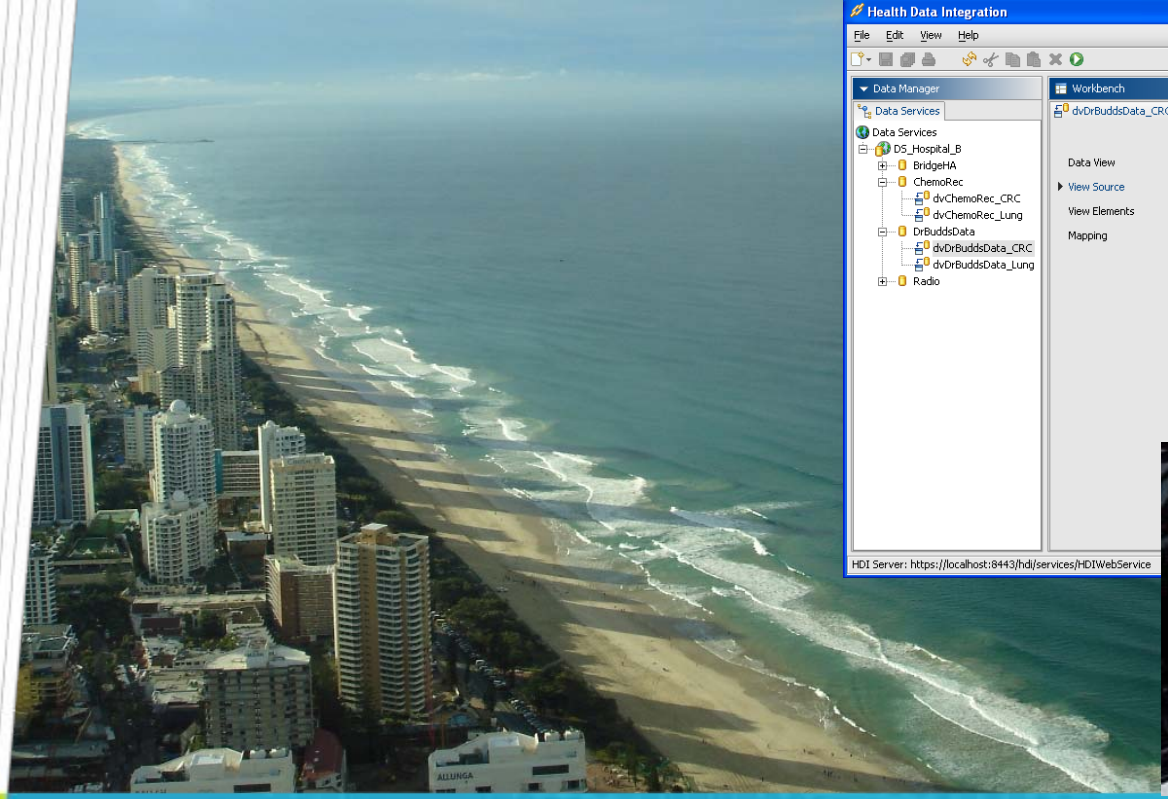
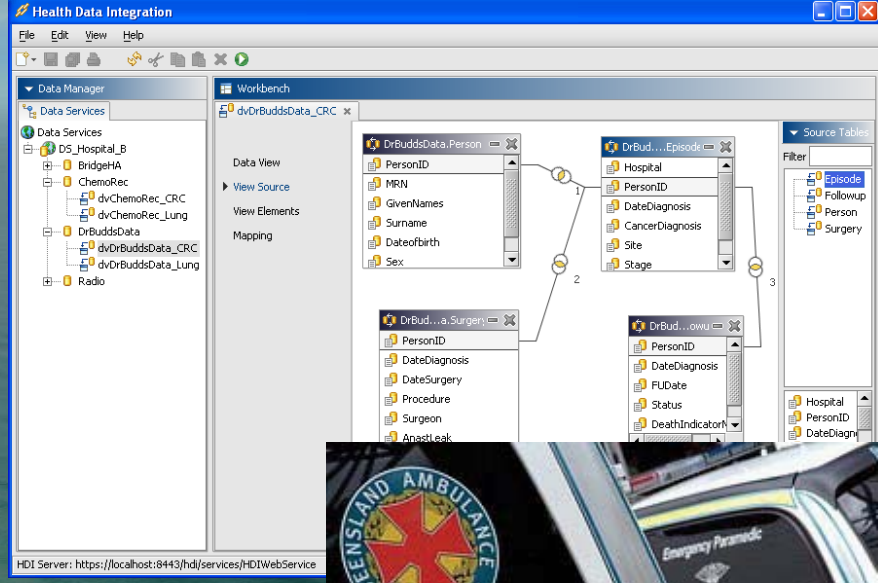




www.csiro.au

Applied Data Linkage: from dirty data to clean complete patient records in emergency health care

Dr Julia Crilly, John O'Dwyer, Marilla O'Dwyer, Dr James Lind, Kerri Melki, Dr Julia Peters, Assoc Prof Vivienne Tippett, Nerolie Bost, Prof Marianne Wallis, Dr Gerben Keijzers

THE AUSTRALIAN E-HEALTH RESEARCH CENTRE

Queensland Government Queensland Health

QUEENSLAND EMERGENCY MEDICINE RESEARCH FOUNDATION

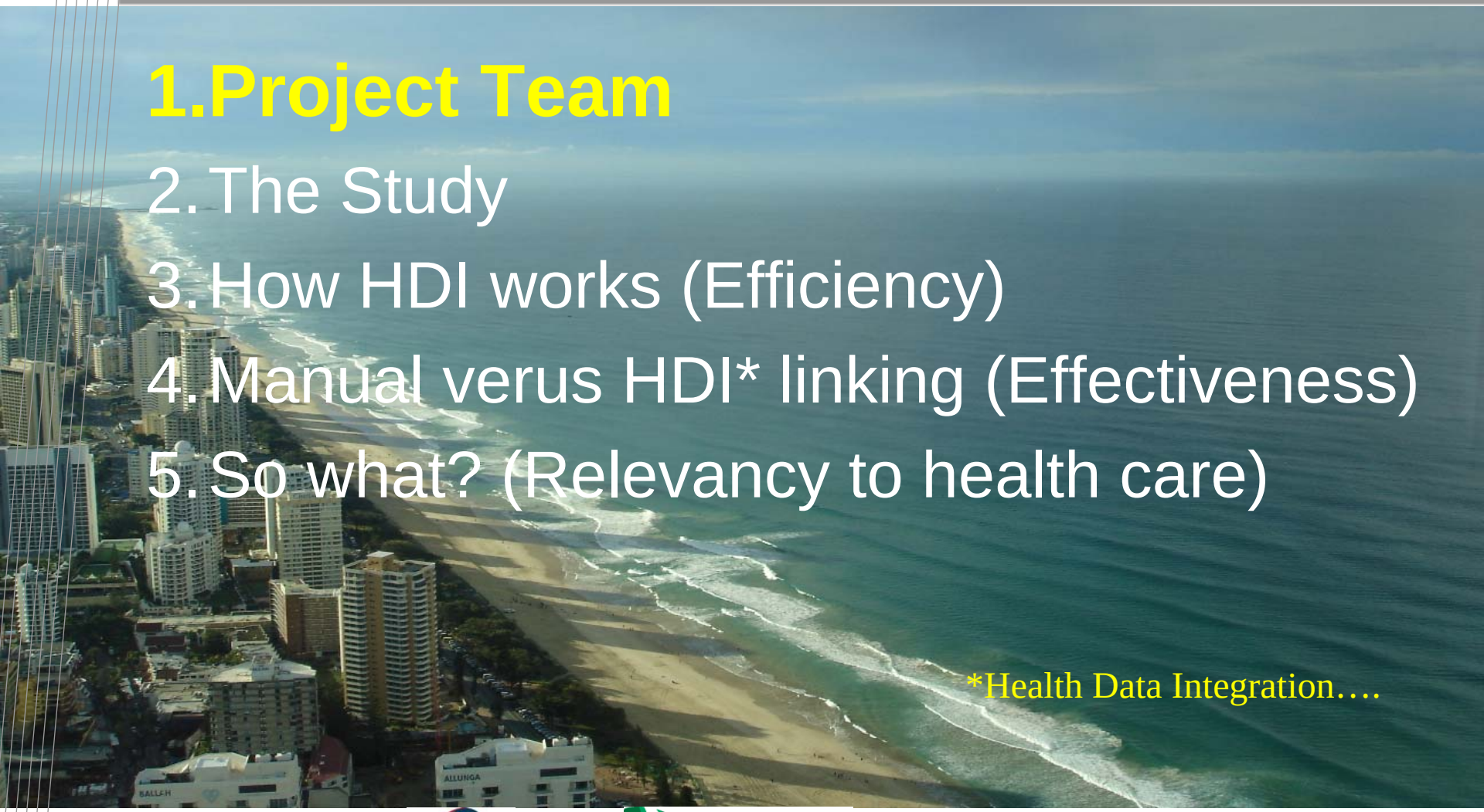
Griffith UNIVERSITY

Australian Centre for Prehospital Research

THE UNIVERSITY OF QUEENSLAND

CSIRO

Overview



1. **Project Team**
2. The Study
3. How HDI works (Efficiency)
4. Manual verus HDI* linking (Effectiveness)
5. So what? (Relevancy to health care)

*Health Data Integration….

THE AUSTRALIAN E-HEALTH RESEARCH CENTRE

Queensland Government Queensland Health

QUEENSLAND EMERGENCY MEDICINE RESEARCH FOUNDATION

Griffith UNIVERSITY

Australian Centre for Prehospital Research

THE UNIVERSITY OF QUEENSLAND

CSIRO

Robina ED Impact (REDI) Project team

Principal Investigators – **Dr James Lind** (Director of Emergency Medicine Training, GCH) **Dr Julia Crilly** (ED Nurse Researcher, GCH, Griffith University)
Co-investigators **Dr Julia Peters** (Registrar, GCH ED) **Nerolie Bost** (Research Officer, GCH ED) **Dr Gerben Keijzers** (Staff Specialist GCH ED, Bond University) **Prof Marianne Wallis** (Nursing Research Chair, GCH & Griffith University)
Marilla O'Dwyer, John O'Dwyer, Kerri Melki (Australian E-Health Research Centre)
Assoc Prof Vivienne Tippett (Director, Australian Centre for Pre-Hospital Research, Queensland Ambulance Service)

GCH: Gold Coast Hospital; ED: emergency department

Contributing Bodies

1. Queensland Health
2. Gold Coast Hospital Foundation
3. Queensland Ambulance Service
4. Australian E-Health Research Centre
5. Australian Centre for Pre-hospital Research
6. Griffith University
7. Queensland Emergency Medicine Research Foundation

Queensland Government

Queensland Health

THE AUSTRALIAN E-HEALTH RESEARCH CENTRE

QUEENSLAND EMERGENCY MEDICINE RESEARCH FOUNDATION

Griffith UNIVERSITY

Australian Centre for Prehospital Research

QUEENSLAND AMBULANCE SERVICE

THE UNIVERSITY OF QUEENSLAND

The Study

1. Project Team
2. **The study**
3. How HDI works (Efficiency)
4. Manual vs HDI* linking (Effectiveness)
5. So what? (Relevance to health care)

*Health Data Integration….

Southern Area Health Service District

LOGAN HOSPITAL

QUEENSLAND AMBULANCE SERVICE

SOUTHPORT HOSPITAL

ROBINA HOSPITAL

The Study

- AIM: to determine the impact of opening a new emergency department in Robina on the patient outcomes and service delivery for the Gold Coast Hospital

- It is suspected that the load balancing will improve
  - Ambulance delivery times
  - Overflow capability and
  - Patient outcomes

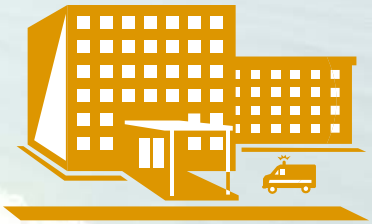
- Required data that can accurately trace each patient journey from Ambulance to ED to Admissions.

- Working with health data is challenging eg. guessed age, partial names (DIRTY DATA)

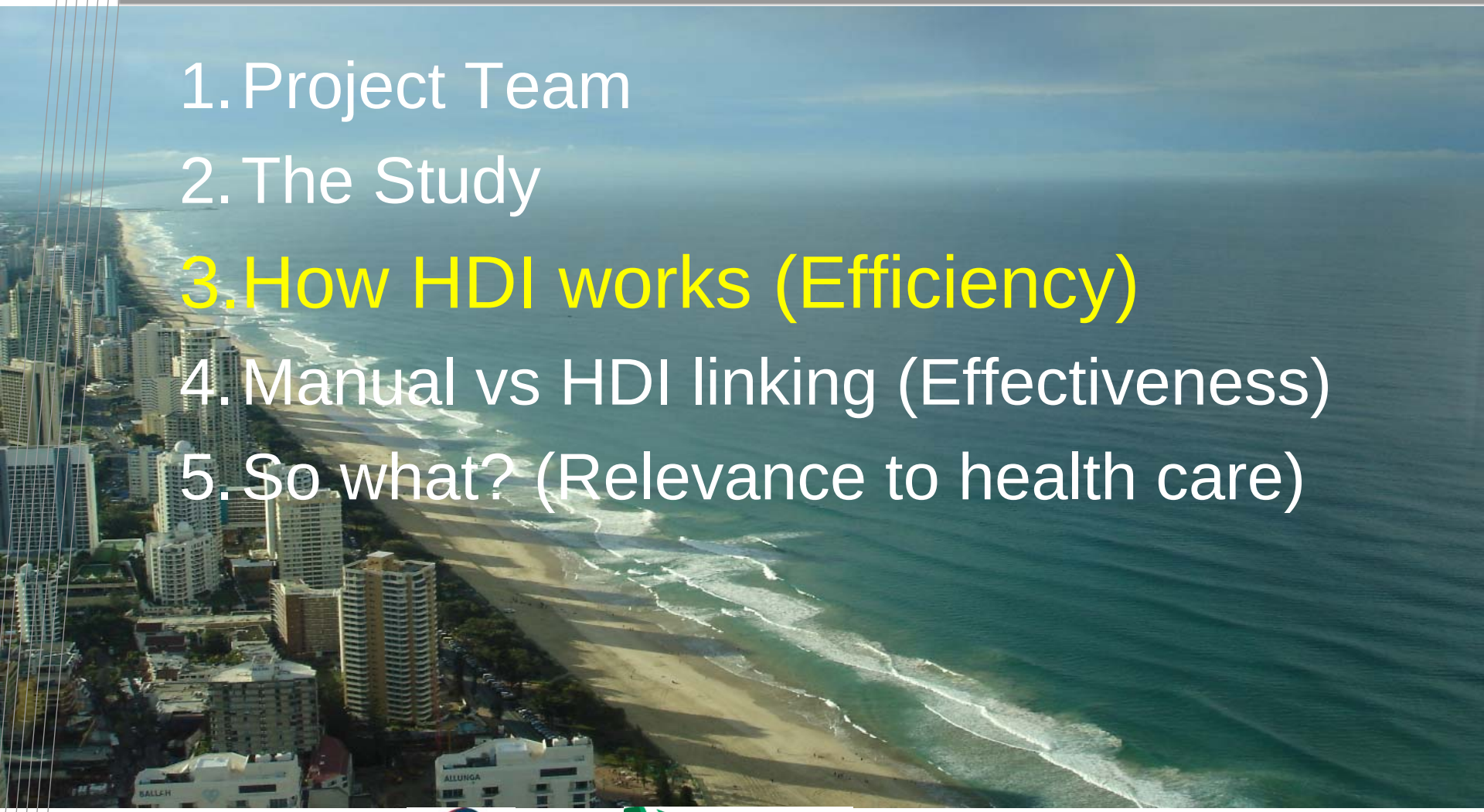
The Study

- Design: One month pre and post comparative study
- Data collection: Using routinely collected health data from three different health information systems (Ambulance, ED, hospital)
- Link three HIS to create one complete data set


+


+


Overview

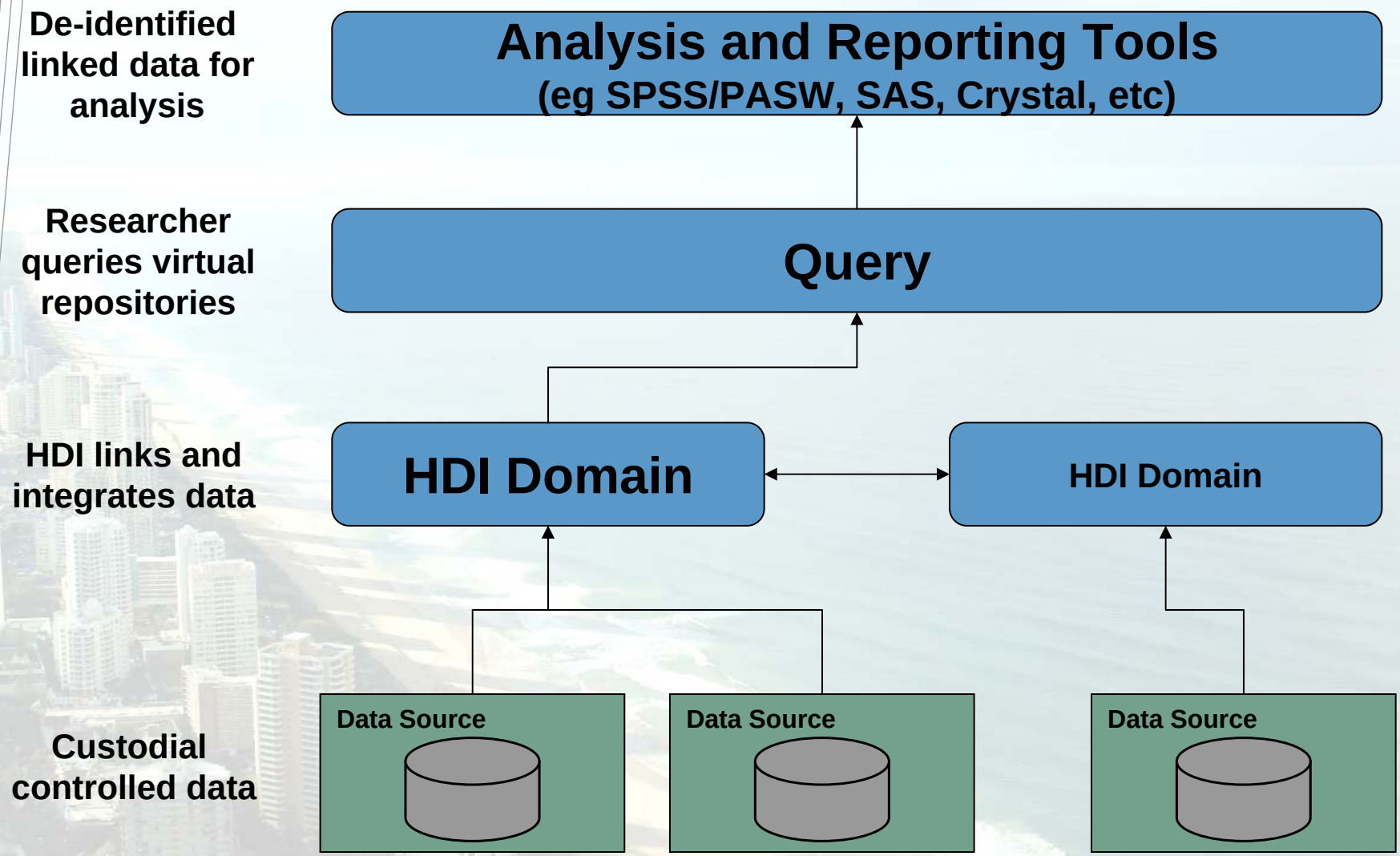


1. Project Team
2. The Study
3. How HDI works (Efficiency)
4. Manual vs HDI linking (Effectiveness)
5. So what? (Relevance to health care)

What is Health Data Integration (HDI)?

- Facilitates linking of patient information across multiple databases, at multiple locations, while maintaining their privacy and security
  - Encrypted linking to protect identity and personal information
  - Federated approach to integrate local data repositories – i.e. many local data sources operating as one and accessed remotely and securely
  - Allows data custodians to retain control of their data - they specify how much of their locally stored data is available to a specific project and in what format
  - Allows researchers or clinicians to access linked but de-identified data sets for further analysis and reporting (as per ethics agreements)

HDI Platform Technology



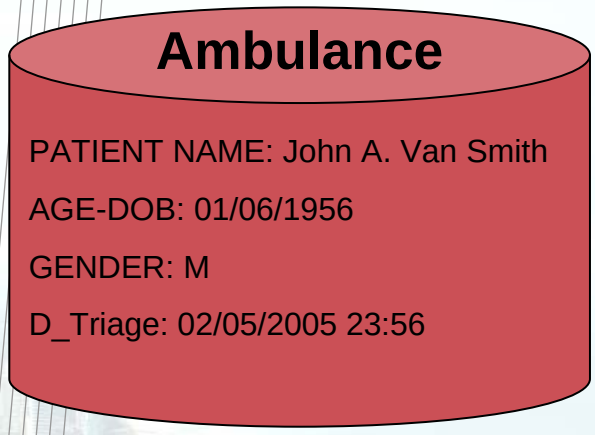
A different approach to matching…..

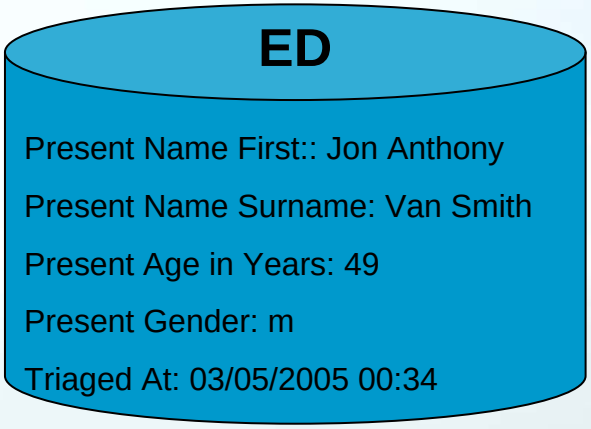
- Rule based approach
- Based on easily understandable ‘human’ approach to matching (Hansen & Maeder, 2007)
- User defined and completely configurable (eg can include ‘case matching’ in the criteria)

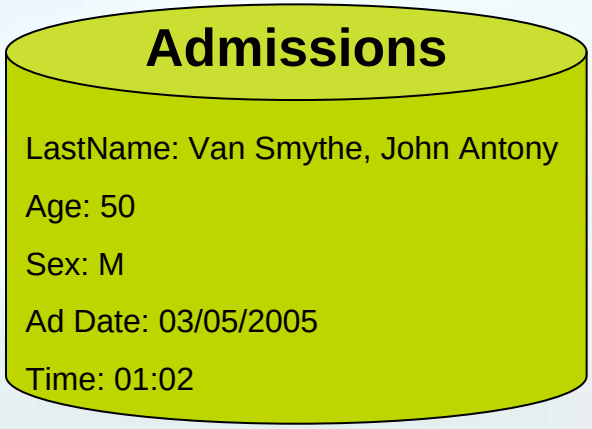
	Probabilistic	HDI ‘rule based’
John Smith, 44, 1, Mon 3.00pm Jon Smyth, 43, 1, Mon 3.35pm	58%	Yes*
Andrea O’Conner, 23, 2, Fri 5.25am Andrea M. O’Connor 23, 1, Fri 6.00am	91%	Yes*
Rebecca Andersen, 38, 1, Tue 10.38pm Rebecca Andersen, 38, 1, Fri 4.52am	99%	No*
Bob Calender, 64, 1, Wed 7.49pm Robert Calendar, 66, 1, Wed 8.15pm	89%	Yes*

* Based on linking algorithms used in REDI project

What the data looks like……

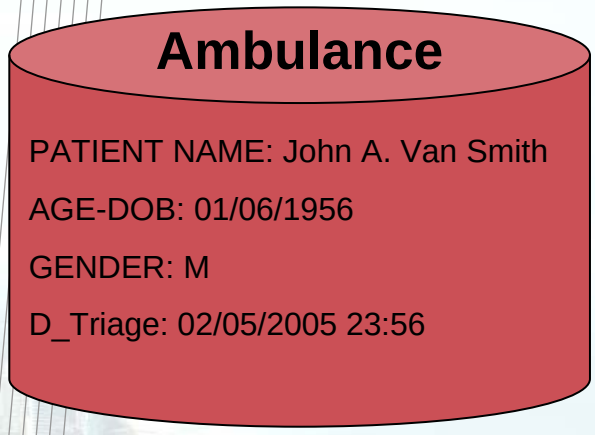


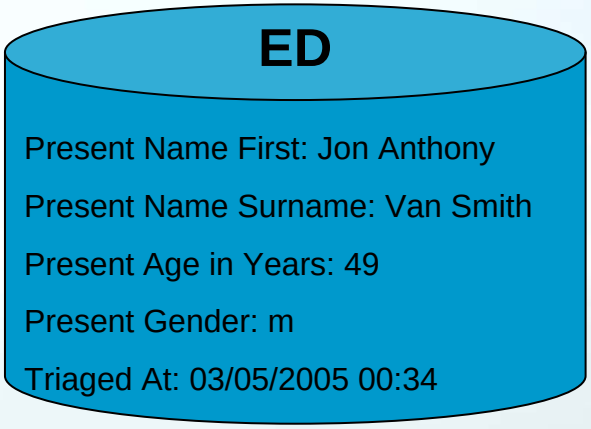


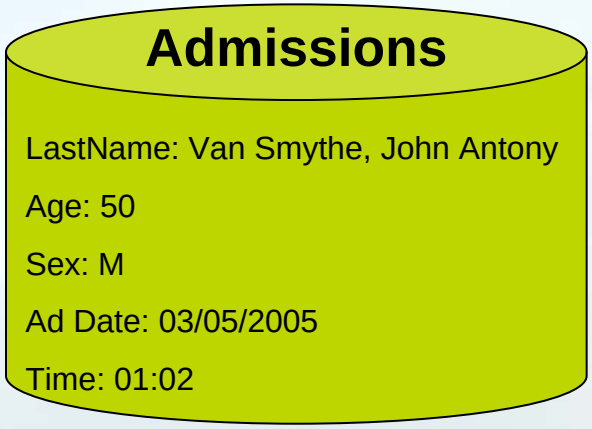


- Names - different formats, spelling
- Age – date of birth, age, estimated age
  - Ie age difference for the same person up to ± 5 years
- Gender – different formats
- Case match – hours between Ambulance – ED and ED-Admissions
- Plus all the usual mis-spellings, data capture / entry errors and typos

How we set it up to match…..







Standardise and transform data……

FirstName: John	FirstName: Jon	FirstName: John
Surname1: Smith	Surname1: Van Smith	Surname1: Van Smythe
Surname2: A. Van Smith	Surname2: Van Smith	Surname2: Van Smythe
Surname3: Van Smith	Surname3: Van Smith	Surname3: Van Smythe
Surname4: Van Smith	Surname4: Van Smith	Surname4: Van Smythe
AgeRange: 44-54	AgeRange: 44-54	AgeRange: 45-55
Sex: 1	Sex: 1	Sex: 1
CaseTimeFrom: 02/05/2005 23:00	CaseTimeFrom: 03/05/2005 00:00	CaseTimeFrom: 03/05/2005 01:00
CaseTimeTo: 03/05/2005 02:00	CaseTimeTo: 03/05/2005 03:00	CaseTimeTo: 03/05/2005 04:00

How it matches…….

Ambulance

PATIENT NAME: John A. Van Smith

AGE-DOB: 01/06/1956

GENDER: M

D_Triage: 02/05/2005 23:56

ED

Present Name First:: Jon Anthony

Present Name Surname: Van Smith

Present Age in Years: 49

Present Gender: m

Triaged At: 03/05/2005 00:34

Admissions

LastName: Van Smythe, John Antony

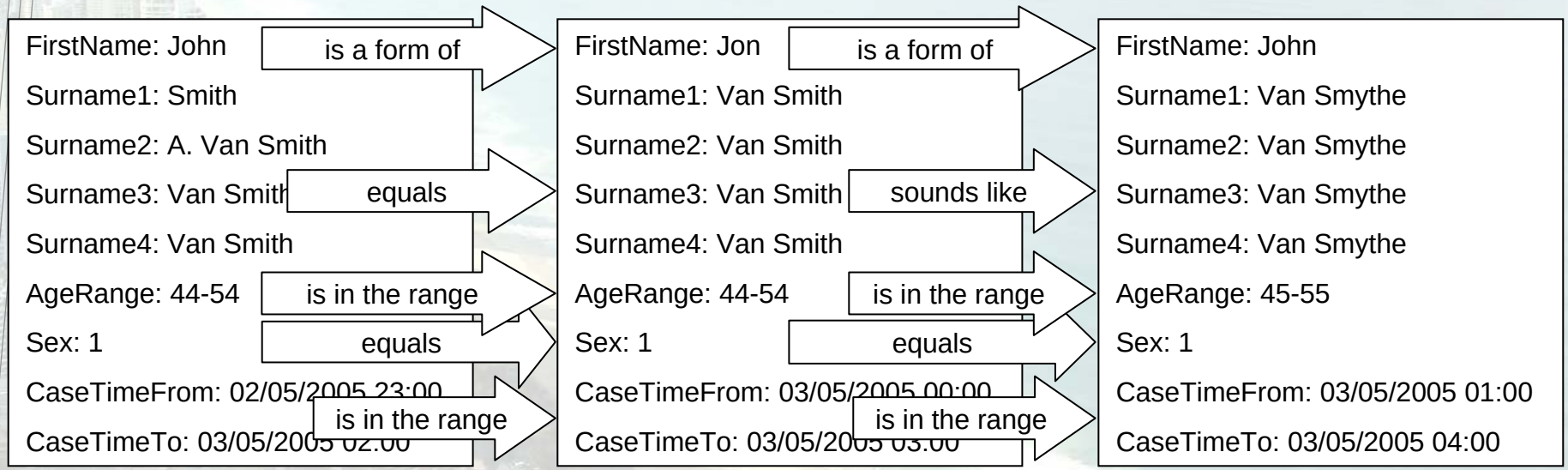
Age: 50

Sex: M

Ad Date: 03/05/2005

Time: 01:02

Link in HDI…… based on user-defined rules…….



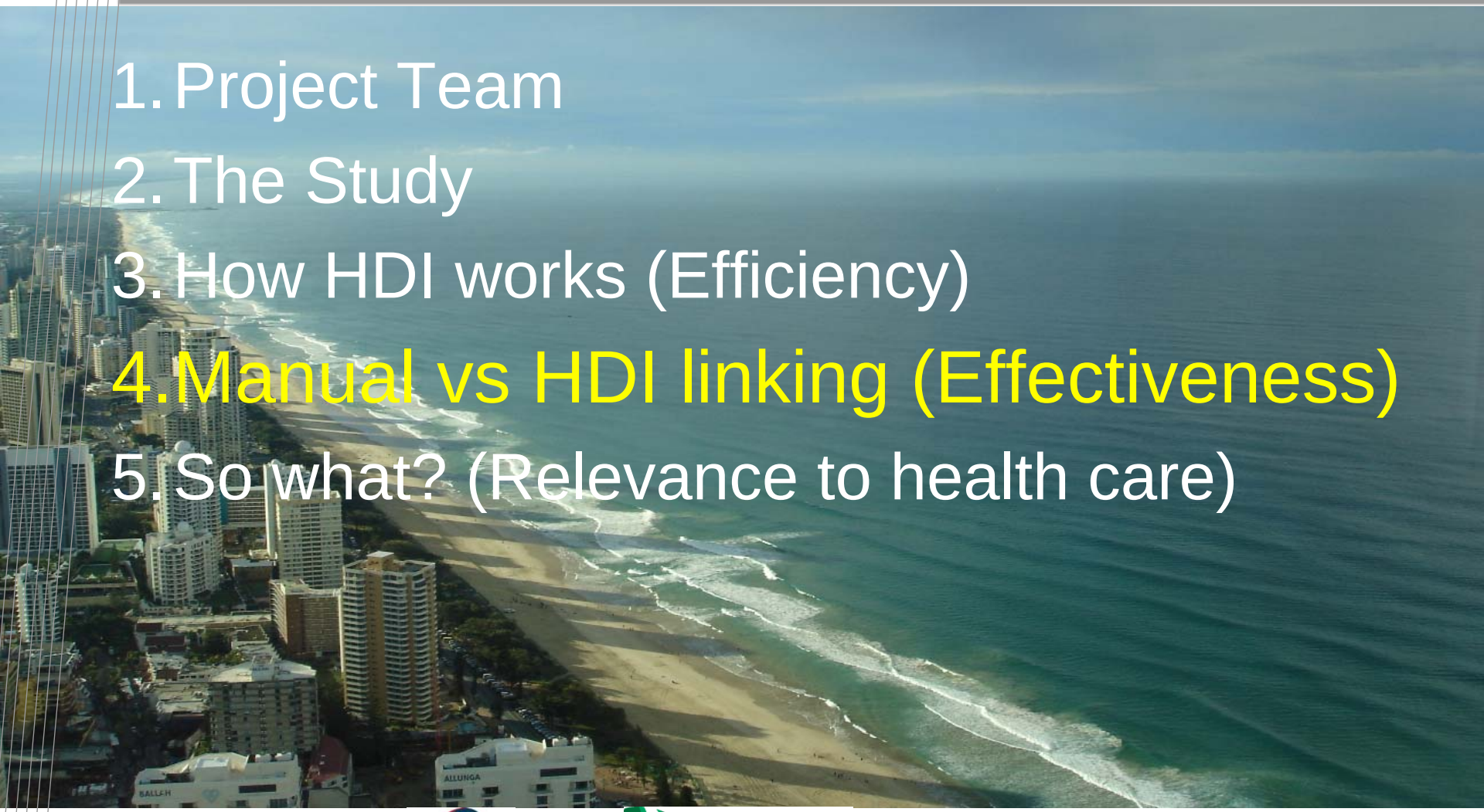
The Challenges

- **While working with the data we had to overcome some unique challenges**

- **Identity matching**
  - Often the age of an individual must be guessed
    - This may not exactly match the age captured in the ED or admissions data
  - Names are often misheard and may only sound like the actual name
    - Like in our example “Smythe” may be captured as “Smith” in the ambulance data

- **Case matching**
  - Match the Ambulance data with a specific Emergency Department arrival
    - Must match person’s ID
      - **AND**
    - The triage time in ED must be within a few hours of the Ambulance triage time
  - Similar measures must be taken with the matching from ED to Hospital Admissions

Overview



1. Project Team
2. The Study
3. How HDI works (Efficiency)
4. Manual vs HDI linking (Effectiveness)
5. So what? (Relevance to health care)

THE AUSTRALIAN E-HEALTH RESEARCH CENTRE

Queensland Government Queensland Health

QUEENSLAND EMERGENCY MEDICINE RESEARCH FOUNDATION

Griffith UNIVERSITY

Australian Centre for Prehospital Research

THE UNIVERSITY OF QUEENSLAND

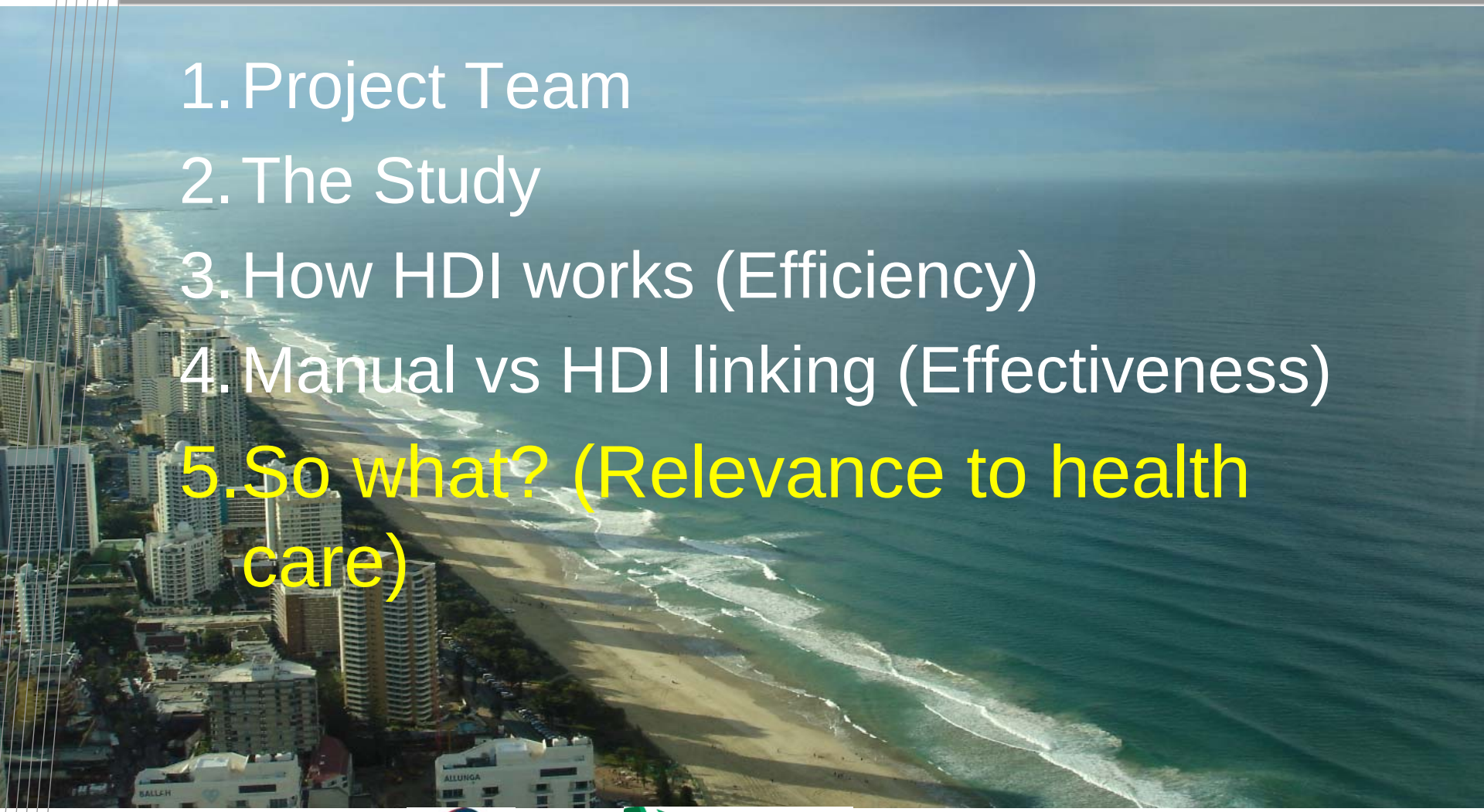
CSIRO

Linking Results: Manual vs HDI

Gold Standard Results (Manual)		
Case		EDIS
	Rows	10,835
QAS	3,469	3,192 (92.0%)
HBCIS	3,431	3,244 (94.5%)

HDI Results		
Case		EDIS
	Rows	10,835
QAS	3,469	3,049 (87.9%)
HBCIS	3,431	3,240 (94.4%)

Overview



1. Project Team
2. The Study
3. How HDI works (Efficiency)
4. Manual vs HDI linking (Effectiveness)
5. So what? (Relevance to health care)

Study Findings

- **10,835 patient presentations were made to the ED during two months**
- **Offload time decreased sig (11 mins to 10 mins, $p < 0.001$)**
- **Proportion offload delay (>15 mins) decreased sig (38% to 27%, $p < 0.001$)**
- **Hosp LOS decreased sig**
- **ED LOS > 8 hrs about 19% in both groups (NS)**

Study Conclusion

The opening of a new ED had a positive effect on ambulance access to the ED, time to unloading of patients and hospital length of stay but had no effect on general ED functioning

(Peters J, Crilly J, Lind J, Tippett V, O'Dwyer M, O'Dwyer J, Melki K, Bost N, Wallis M, Keijzers G 2008)

File Edit View Data Transform Analyze Graphs Utilities Window Help

1:

Visible: 44 of 44 Variables

	q_AgeAtDiagnosis	q_Sex	q_C_DISPATCH_CRITI	q_T_FINAL_ASSESSMENT	q_D_DISPATCH	q_D_AT_SCENE	q_D_AT_DEST	q_D_TRIAGE	q_D_OFF_STRETCHER	q_D_COMPLETED	e_mrn_medical_record_number	e_AgeAtDiagnosis	e_Sex	e_present_postcode	e_diagnosis_icd_code	e_mode	e_triage_priority	
1	45	2	2	FINAL ASSESSMENT UNKNOWN	2007-01-22 11:54	2007-01-22 12:06	2007-01-22 12:33:0	2007-01-22 12:35	2007-01-22 12:45	2007-01-22 12:56:0	345698	45.0	2	4212	N39.0	AR	2	2007-
2	.	.	.								456222	55.0	2	3111	T41.3	W	3	2007-
3	.	.	.								888234	43.0	2	4209	K59.0	AR	3	2006-
4	30	2	1	DRUG INTOXICATION	2006-09-17 01:29	2006-09-17 01:46	2006-09-17 02:19:0	2006-09-17 02:28	2006-09-17 02:28	2006-09-17 06:28:0	789436	30.0	2	4218		AR	3	2006-
5	.	.	.								234789	67.0	2	4211	O72.1	AR	3	2006-
6	.	.	.								674508	57.0	2	2462	K50.9	W	3	2008-
7	.	.	.								812945	43.0	1	4662	Z53.2	W	4	2008-
8	.	.	.								874523	32.0	1	4573	R11	W	3	2008-
9	.	.	.								670945	47.0	2	4216		W	4	2007-
10	.	.	.								777834	4.0	1	4227	L02.8	W	3	2007-
11	.	.	.								569123	.0	2	4214	J06.9	W	3	2008-
12	.	.	.								398127	26.0	2	4211	J18.1	W	3	2007-
13																		

1:

Visible: 44 of 44 Variables

	e_diagnosis_icd_code	e_mode	e_triage_priority	e_arrival_date	e_triaged_at	e_treating_doctor_seen_at	e_departure_ready_at	e_departure_actual_at	e_departure_st	e_departure_status_desc
1	N39.0	AR	2	2007-01-22 12:43:00.000	2007-01-22 12:44:00.000	2007-01-22 12:55:00.000	2007-01-22 18:00:00.000	2007-01-22 18:24:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED
2	T41.3	W	3	2007-09-02 10:22:00.000	2007-09-02 10:22:00.000	2007-09-02 10:27:00.000	2007-09-02 16:20:00.000	2007-09-02 16:20:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED
3	K59.0	AR	3	2006-09-08 03:02:00.000	2006-09-08 03:05:00.000	2006-09-08 03:24:00.000	2006-09-08 06:43:00.000	2006-09-08 06:43:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED
4		AR	3	2006-09-17 02:24:00.000	2006-09-17 02:24:00.000		2006-09-17 02:55:00.000	2006-09-17 02:55:00.000	W	DID NOT WAIT
5	O72.1	AR	3	2006-11-17 01:07:00.000	2006-11-17 01:07:00.000	2006-11-17 01:28:00.000	2006-11-17 03:19:00.000	2006-11-17 08:02:00.000	A	ADMITTED (EXCL. ED BED)
6	K50.9	W	3	2008-05-29 14:48:00.000	2008-05-29 14:48:00.000	2008-05-29 15:56:00.000	2008-05-29 18:38:00.000	2008-05-29 20:12:00.000	A	ADMITTED (EXCL. ED BED)
7	Z53.2	W	4	2008-03-10 18:34:00.000	2008-03-10 18:38:00.000		2008-03-10 23:25:00.000	2008-03-10 23:25:00.000	W	DID NOT WAIT
8	R11	W	3	2008-05-04 12:04:00.000	2008-05-04 12:05:00.000	2008-05-04 13:46:00.000	2008-05-04 17:13:00.000	2008-05-04 17:13:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED
9		W	4	2007-04-10 21:06:00.000	2007-04-10 21:06:00.000	2007-04-10 22:48:00.000	2007-04-11 00:59:00.000	2007-04-11 00:59:00.000	W	DID NOT WAIT
10	L02.8	W	3	2007-01-22 16:29:00.000	2007-01-22 16:29:00.000	2007-01-22 19:35:00.000	2007-01-22 21:49:00.000	2007-01-23 01:46:00.000	A	ADMITTED (EXCL. ED BED)
11	J06.9	W	3	2008-04-11 10:00:00.000	2008-04-11 10:00:00.000	2008-04-11 10:23:00.000	2008-04-11 10:58:00.000	2008-04-11 11:00:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED
12	J18.1	W	3	2007-06-08 22:11:00.000	2007-06-08 22:21:00.000	2007-06-09 00:23:00.000	2007-06-09 02:45:00.000	2007-06-09 05:10:00.000	A	ADMITTED (EXCL. ED BED)
13										

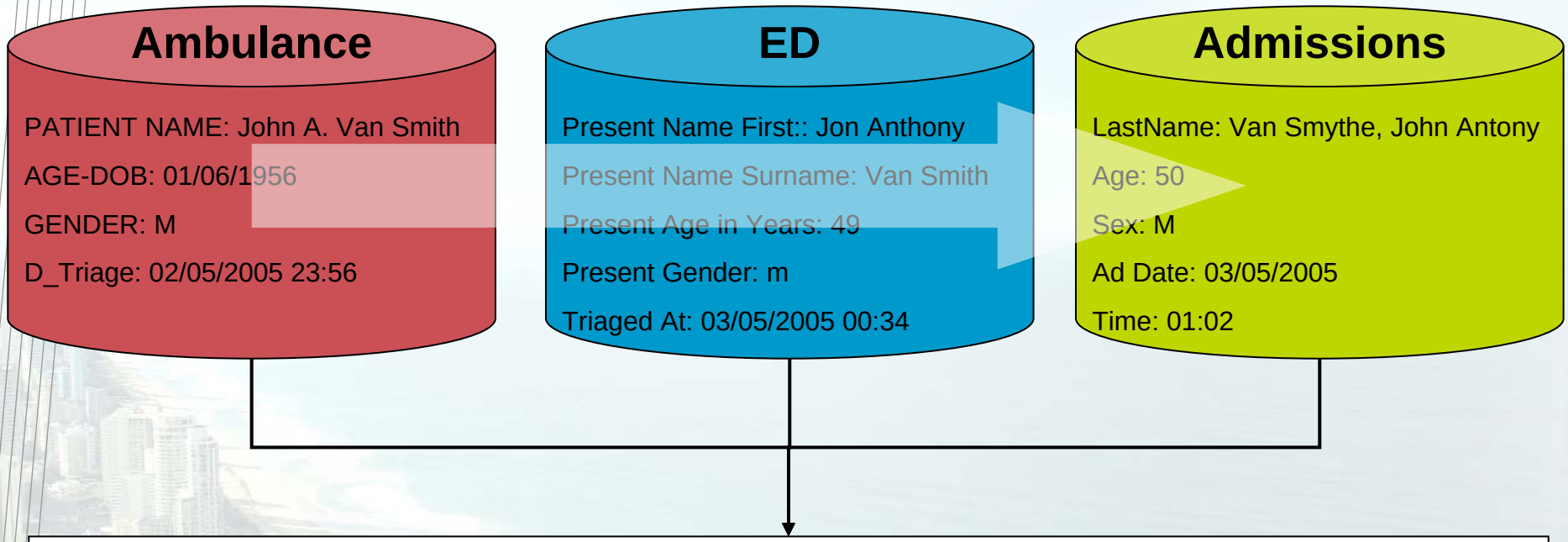
File Edit View Data Transform Analyze Graphs Utilities Window Help

29 : h_los

Visible: 44 of 44 Variables

	e_departure_actual_at	e_departure_status	e_departure_status_desc	e_occurances	h_medical_record_number	h_enc_no	h_Sex	h_AgeAtDiagnosis	h_post_code	h_ad_source	h_ad_datetime	h_trg_category	h_care_type	h_ad_unit	h_ad_ward		
1	2007-01-22 18:24:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED	1	345698	345698-1	2	45.0	4225	2	2007-01-22 18:00	2	1	MED	ED	4.0	2007-01-2
2	2007-09-02 16:20:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED	1	.		.	.	.	.		.	.			.	
3	2006-09-08 06:43:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED	1	.		.	.	.	.		.	.			.	
4	2006-09-17 02:55:00.000	W	DID NOT WAIT	1	.		.	.	.	.		.	.			.	
5	2006-11-17 08:02:00.000	A	ADMITTED (EXCL. ED BED)	1	.		.	.	.	.		.	.			.	
6	2008-05-29 20:12:00.000	A	ADMITTED (EXCL. ED BED)	1	674508	674508-4	2	57.0	2462	3	2008-05-29 18:33	2	1	GYN	ED	2.0	2008-05-3
7	2008-03-10 23:25:00.000	W	DID NOT WAIT	1	.		.	.	.	.		.	.			.	
8	2008-05-04 17:13:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED	1	.		.	.	.	.		.	.			.	
9	2007-04-11 00:59:00.000	W	DID NOT WAIT	1	.		.	.	.	.		.	.			.	
10	2007-01-23 01:46:00.000	A	ADMITTED (EXCL. ED BED)	1	777834	777834-2	1	4.0	4213	3	2007-01-22 21:48	2	1	PAE	ED	4.0	2007-01-2
11	2008-04-11 11:00:00.000	H	ED SERVICE EVENT COMPLETED - DISCHARGED	1	.		.	.	.	.		.	.			.	
12	2007-06-09 05:10:00.000	A	ADMITTED (EXCL. ED BED)	1	398127	398127-3	2	26.0	4211	3	2007-06-09 02:40	2	1	MD1	GMAU	1.0	2007-06-0
13																	

The REDI Result: a linked health care journey



John Van Smith, approx 45, Male

Linked data for John allows us to answer research questions:

How long did it take the ambulance to get to John, from the time he called the ambulance?

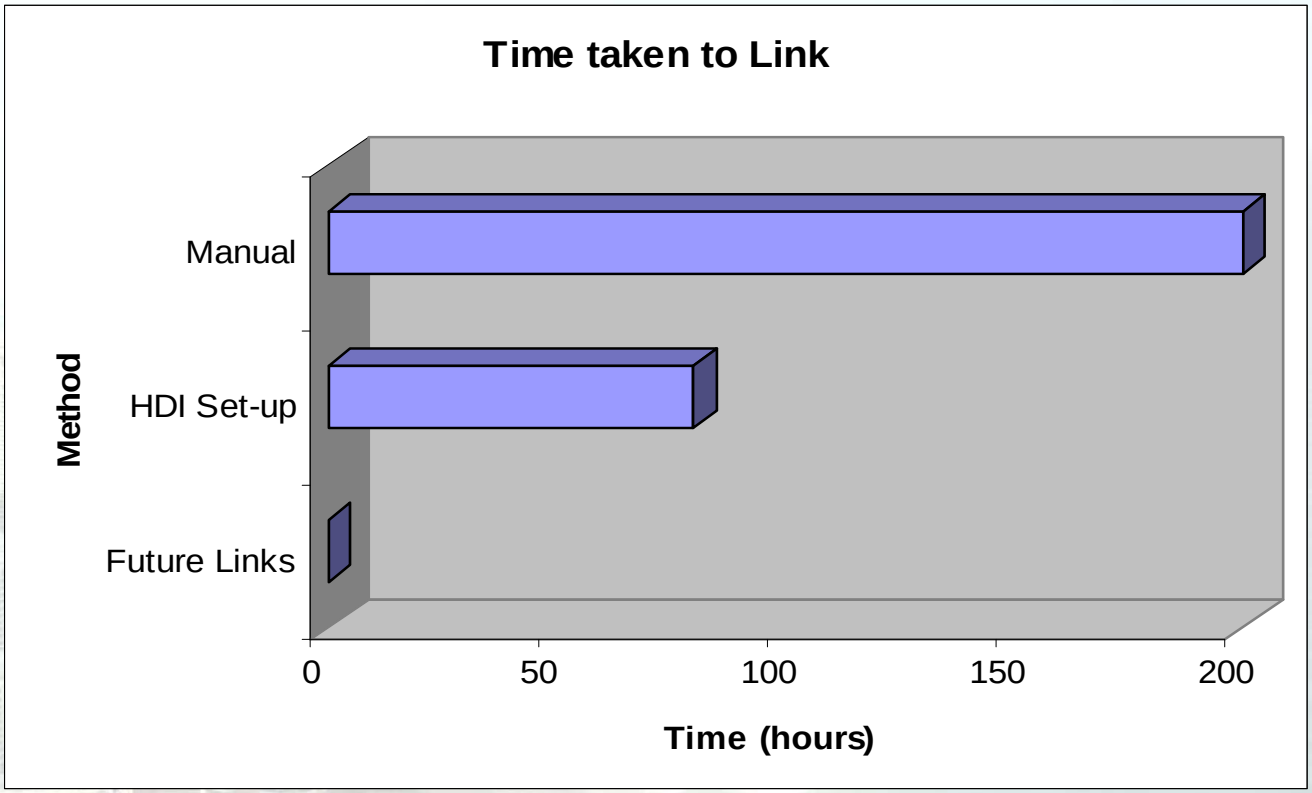
How long did John wait to see a Dr in the ED?

How long did he wait in the ED?

Did he get admitted to hospital? Did he die in hospital?

So what?
(Relevance to health care RESEARCHERS)

1. Time saving



2. Allows researchers to access linked but de-identified data sets for further analysis and reporting (as per ethics agreements)
3. Inform policy and practice development

So what?
(Relevance to health care PROVIDERS)

- Potential to have an integrated health care record
- Utilizing existing systems and information
- More ‘real time’
- Multiple hospitals - benchmark

So what?
(Relevance to health care PROVIDERS)

Government of Western Australia
Department of Health

Delivering a Healthy WA

Healthy Workforce · Healthy Hospitals · Healthy Partnerships · Healthy Communities · Healthy Resources · Healthy Leadership

Search: GO

WA Public Hospital Activity

- Home
- ED locations and contacts
- ED activity - now
- ED daily activity
- ED weekly activity
- Hospital beds
- Health care services
- Publications and reports
- Overview of ED
- Contact us
- Back to Department of Health homepage

Delivering a Healthy WA

» Home » ED activity – now

ED activity – now

The table below provides an up-to-date view of the status of the Emergency Department in each of the metropolitan hospitals that contains an Emergency Department.

Preview of Emergency Department (ED) figures at Tuesday, 11 August 2009 02:24 PM

Hospital	Ambulance diverted	Triage 4 patients average waiting time (minutes)	Number of patients waiting to be seen in ED	Total patients in ED
Armadale/Kelmscott District Memorial Hospital	No	45	12	38
Fremantle Hospital	No	71	3	40
Joondalup Health Campus	No	67	15	53
King Edward Memorial Hospital For Women	No	36	2	2
Princess Margaret Hospital For Children	No	25	7	31
Rockingham General Hospital	No	66	15	31
Royal Perth Hospital	No	63	10	53
Sir Charles Gairdner Hospital	No	145	13	56
Swan District Hospital	No	27	4	17

Print the whole page

wa.gov.au

Home | Disclaimer and Privacy

So what?
(Relevance to health care PATIENTS)

- Improved outcomes and safety
- Benefit from efficiencies in process improvement

Now What…

- Stage 1 - Completion of pilot

(2 months data, 1 hospital)

  - Create one dataset showing all patient presentations to emergency via ambulance, then admitted into hospital.
  - Analyse data

- Stage 2 – Complete research project

(2 years data, 3 hospitals)

  - Setup and configure a production ready HDI Server
  - Link and merge a larger dataset showing all patient presentations to emergency via ambulance, then admitted into hospital.
  - Analyse data

Questions?