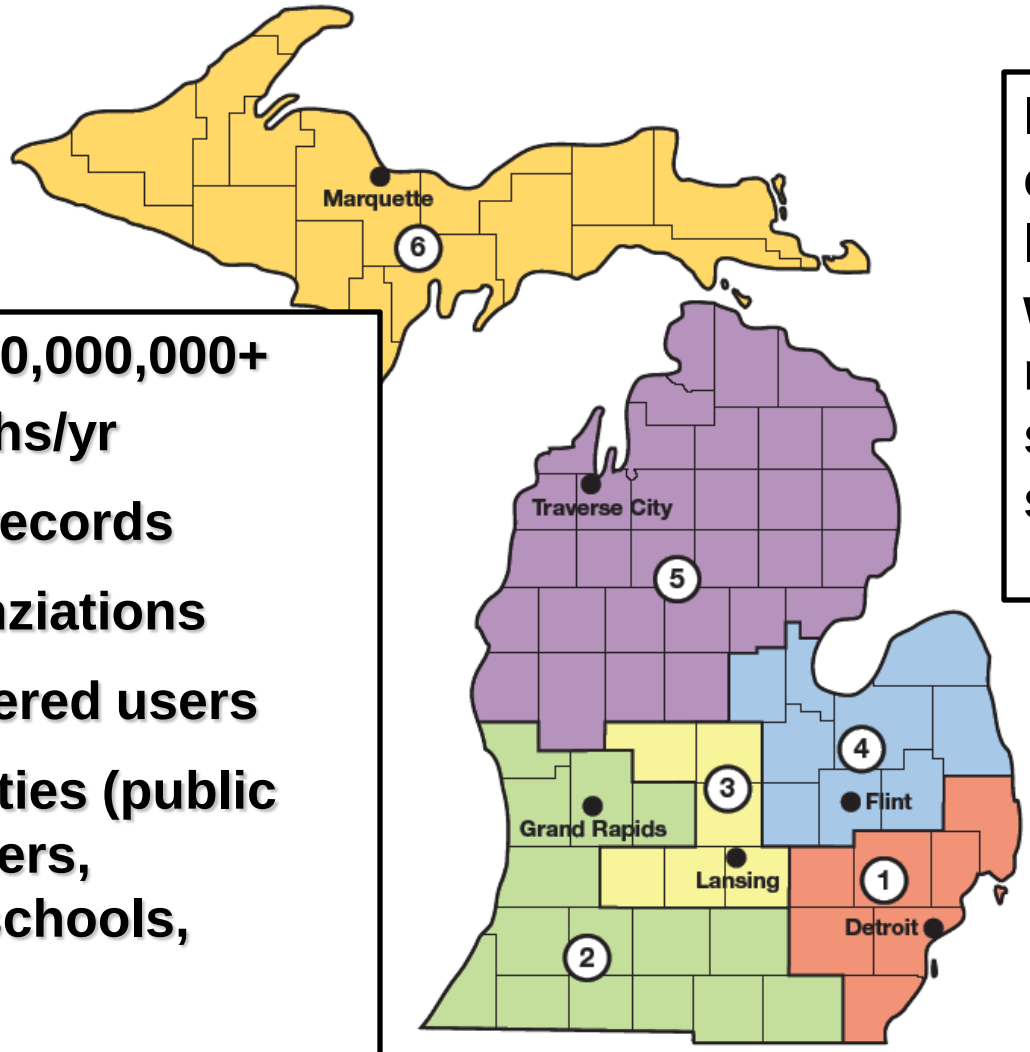
Michigan Care Improvement Registry

Integrating the MCIR with the State of Michigan Enterprise (SOM) HIE Master Person Index (MPI)

October 8, 2013

MCIR Background



Population: 10,000,000+

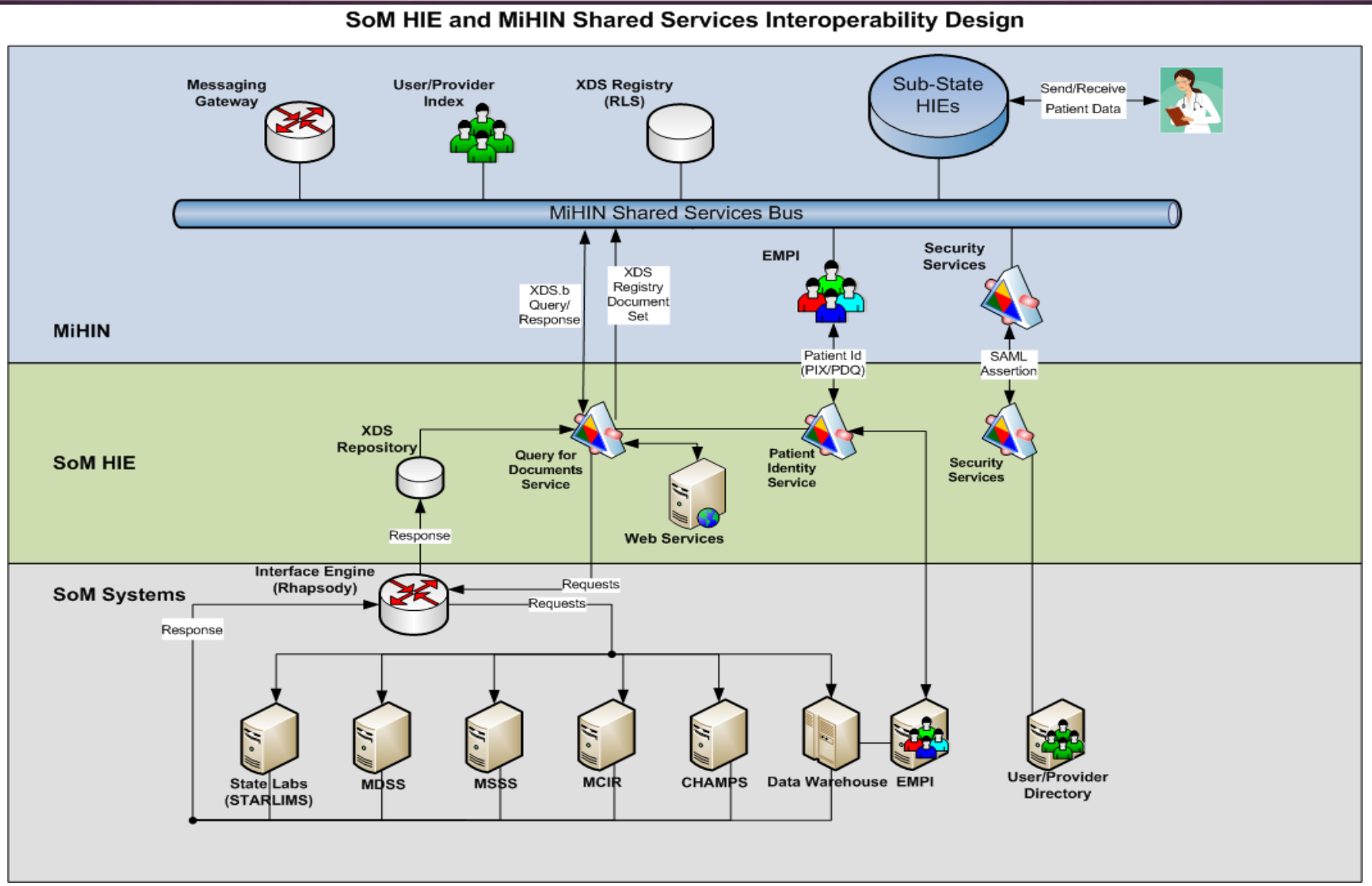
- **120,000+ births/yr**
- **9m+ patient records**
- **90m+ immunziations**
- **25,000 registered users**
- **10,000+ facilities (public health, providers, pharmacies, schools, childcare)**

Interfaces: web client, HL7, legacy transfer, web services for newborn screening, BMI, sickle cell, lead

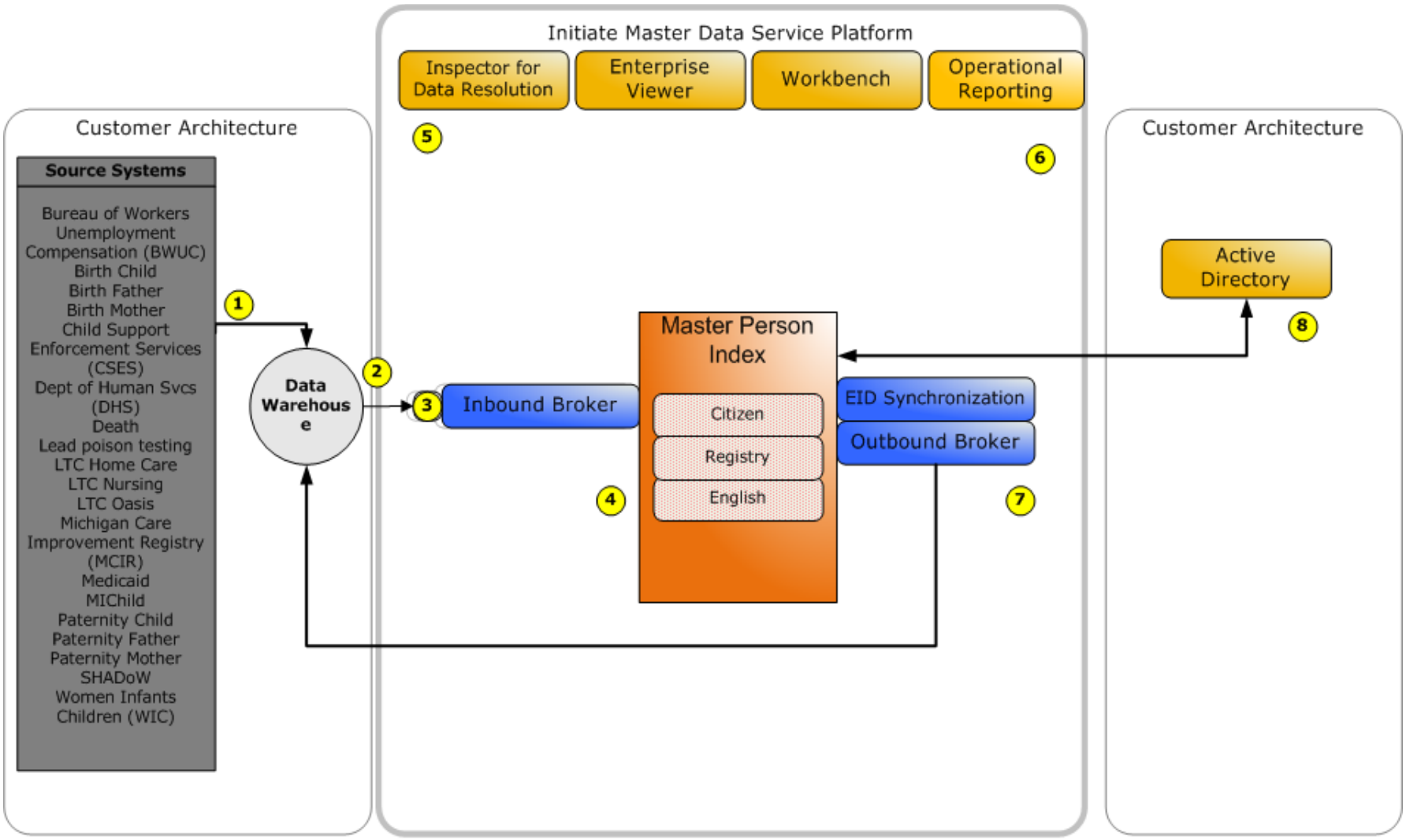
MPI Background

- A “Master Person Index” is designed to store and manage key data from multiple source systems, linking relevant records based on statistical matching algorithms
- Using a sophisticated and flexible data model and attribute weighting, the probabilistic algorithms identify and link records across these source systems
- Enables the combination, comparison, review, and resolution of potential data issues.
- The SOM MPI currently contains 40M records on 14M persons (“Citizens”).

MCIR and SOM Shared Services



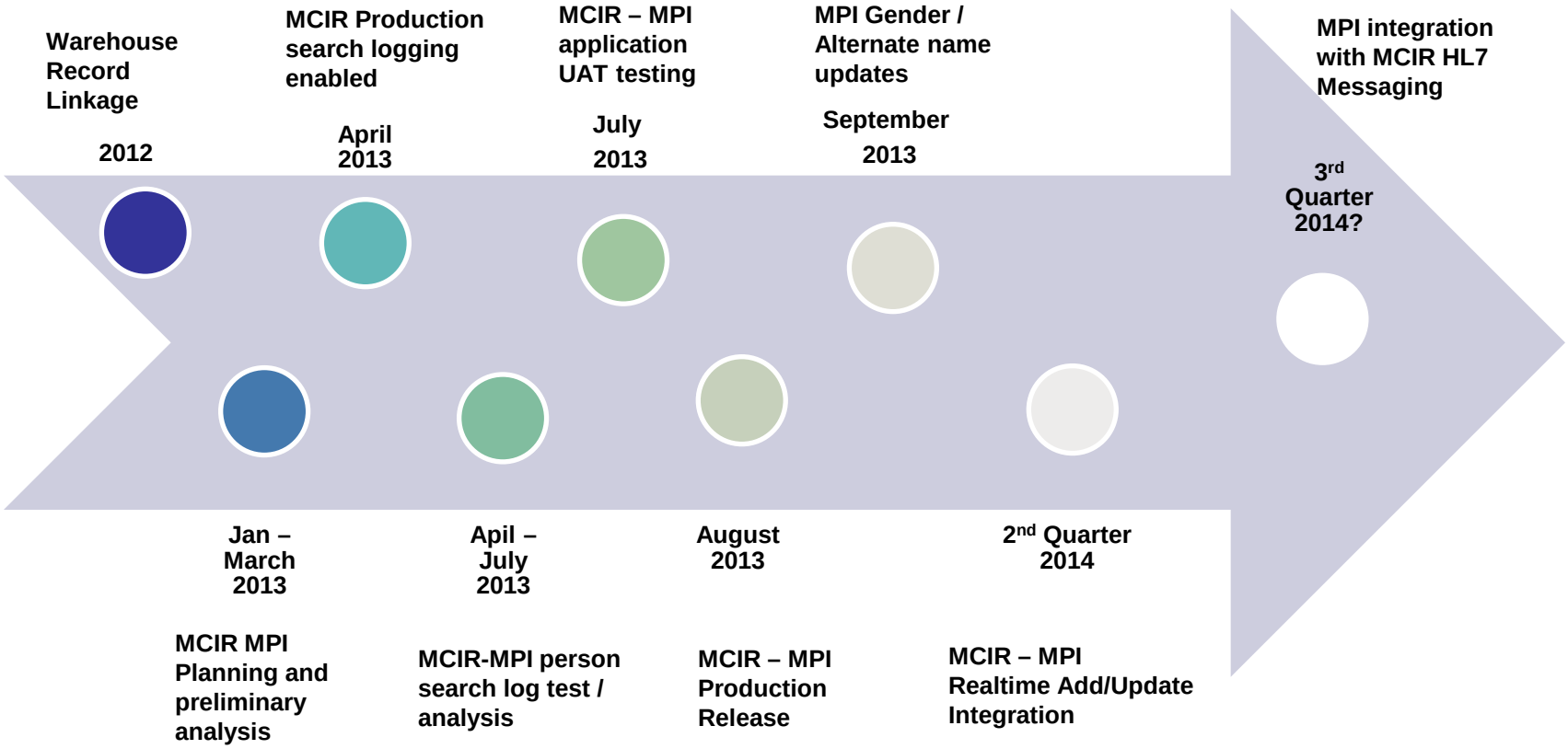
SOM MPI Platform



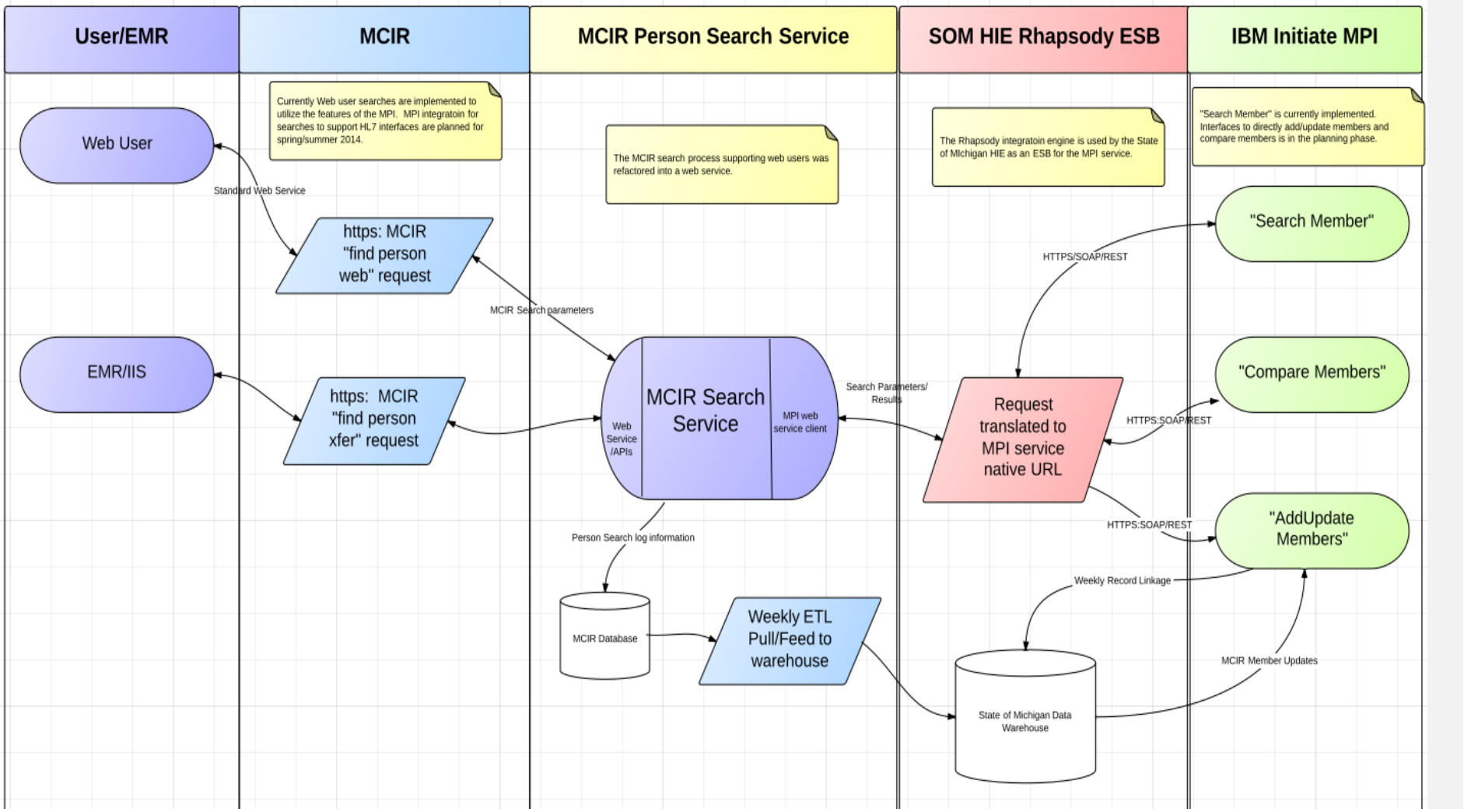
Project Approach for Success

- MCIR and IISs in general have robust requirements for patient search / de-duplication.
- We are a “pea” in the pod and do not own the MPI. An IIS must engage early to influence an MPI implementation so that we are not locked out of accessing required features or interfaces.
- Build community tools as much as possible that can be used for other programs planning to leverage an MPI.
- We must log the IIS production search requests with results and have ability to “replay” search requests against the MPI.
- A transparent “fail-safe” search method to an MPI must be available should the service become unavailable.

MCIR-SOM MPI Timeline



MCIR – MPI Interactions



Necessary Preparatory Steps

- Prioritize candidate interfaces (web client, HL7, Hedis, ..) that are initial targets focusing on the “low hanging fruit”.
- Not all interfaces (web client, HL7, Hedis, etc..) and search types (patient ID/Chart#, family member/responsible party) will be candidates for the MPI. For MCIR Hedis workload is too large to be practical (avg 1-2 million patient record queries per request).
- Analysis of production workload by interface and search type is required to identify appropriate use cases and help the MPI team plan their required infrastructure/scalability.
- Start logging production search workload and results as early as possible to feed your use cases.

Common Toolsets

- MCIR Search log format is based on Google protocol buffer format (http://code.google.com/p/protobuf/).
- A Java client was developed that provides the following features:
  - Read from the production search log
  - Generate CSV or display production search log records
  - Re-submit production search log to MPI using standard web service client used by IBM/Initiate. Other MPIs will require a different web service client.
  - Compare MPI results with IIS search log results
  - Provide summary of analysis / results.
- SAS statistical routines were developed that use the csv export of the production worklog.

Search Logging Tool

- Google protocol buffer format used by MCIR search log is a data interchange format that allows a very efficient encoding of structured data.
- Stored in raw format in database table, consumes minimal space and has very low overhead. Allows record extraction from the raw binary format.
- Logging of following information:
  - Search request parameters
  - Search results
  - Search response time
  - Service providing search results (MPI, MCIR Search service)

Search Log Information

ID: 1317585

Type: PersonSearchLogProtoGen FindPersonScreen

Search timestamp: 20130625000515692 on MCIR system.

Search issued by user ID: Bragg123 for Site ID:

Patient Last Name: Bullwinkle

Patient First Name: Moose

Birth Date (yyyyMMdd): 19381116

Gender (M/F/U): M

MCIR Search Result count: 1

MCIR Search Elapsed Time (ms): 94

MCIR Search returned MCIR_ID: 123456789, Last=Bullwinkle, First=Moose, Birth Date=19381116, Gender=M, RespLastNm=Squirrel, RespFirstNm=Rocky

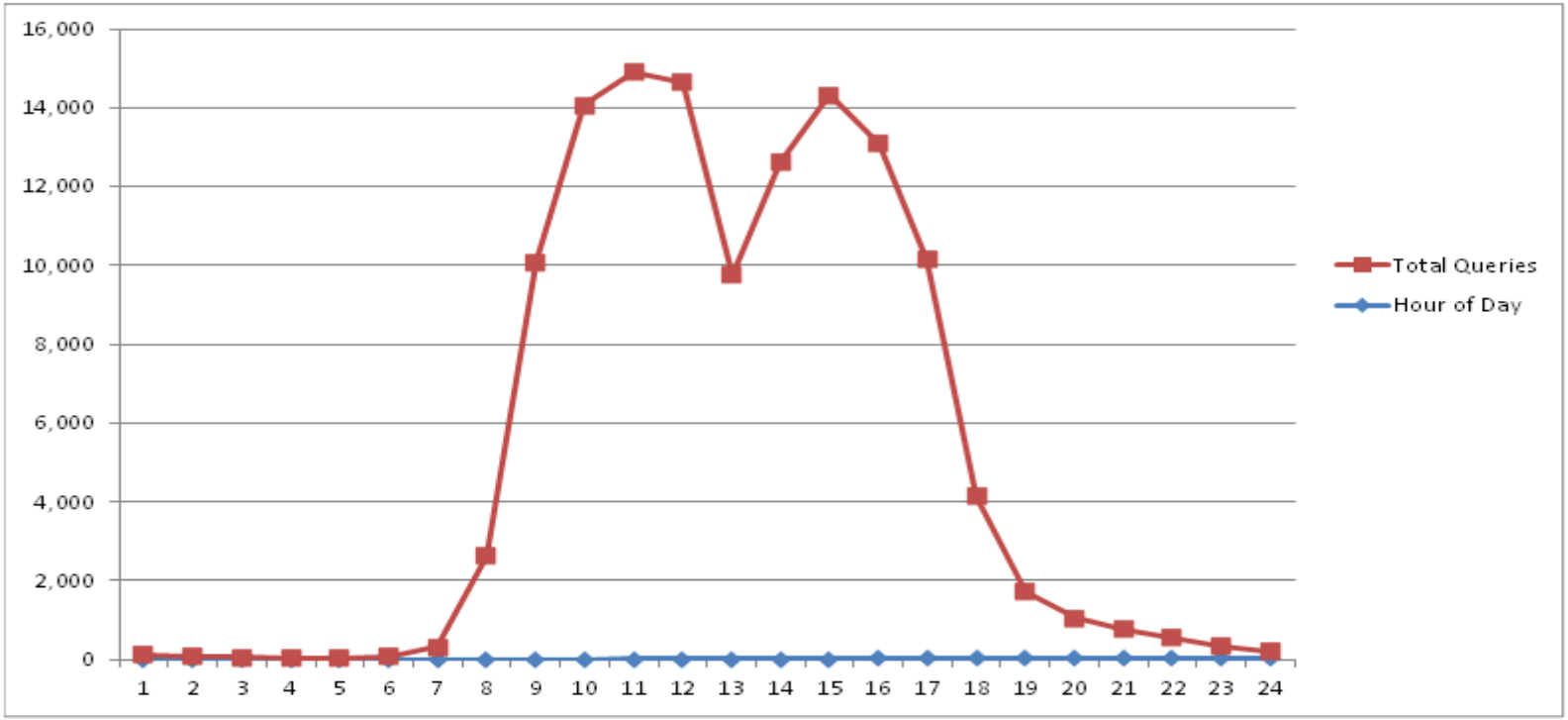
Algorithms count: 0

Distribution of User Searches

MCIR Preliminary Analysis

Mcir ID	First	Last	Birth Date	Gender	Mothers Maiden	Guardian Last	Guardian First	Phone #	WIC ID	Medicaid ID	Plurality	Birth Order	Search Signature	# of queries lookup	% of Total Queries	Avg Response Time (Secs)
0	0	1	1	0	0	0	0	0	0	0	0	0	0011000	131871	55.30%	0.0946
0	1	1	1	0	0	0	0	0	0	0	0	0	0111000	33895	14.21%	0.0744
1	0	0	0	0	0	0	0	0	0	0	0	0	1000000	19101	8.01%	0.0077
0	1	1	1	1	0	0	0	0	0	0	0	0	0111100	13770	5.77%	0.0936
0	1	0	1	0	0	0	0	0	0	0	0	0	0101000	12138	5.09%	0.042
0	2	2	1	0	0	0	0	0	0	0	0	0	0221000	10001	4.19%	0.0969
0	1	1	0	0	0	0	0	0	0	0	0	0	0110000	7637	3.20%	0.0618

Web User Search Workload



The following table shows the variability between peak and average statistics within a given gtime period.
The peak queries per second are magnitudes greater than the maximum # experienced over a minute or hour.
These statistics reflect patient queries performed by end users of the MCIR interactive web application only.

MCIR Interactive Search Workload

		Extrapolated Values for given statistic				
Statistic	Observed	Per Secon	Per Minute	Per Hour	Per Day	Conclusion
Daily Max	125,500	1.45	87.15	5,229.17	--	
Hourly Max	15,000	4.17	250.00	--	360,000	If sustained will be 3 times the observed daily workload
Minute Max	450	7.50	--	27,000.00	648,000.00	If sustained will be 1.8 times the observed hourly workload
Second Max	40	--	2,400	144,000	3,456,000	if sustained will be 6 times the observed per minute workload

Analysis / Test Toolset

- Standalone java client provides:
  - Extracts/downloads search log information
  - Generates CSV of search log / analysis information
  - Re-submits search log workload to MPI web service and generates search log of results
  - Generates comparison of IIS search results / performance with MPI search results / performance for each search attempt.
- Analysis data supported the development/refinement of algorithms used to interpret results from MPI and setting of required scoring threshold.
- Automated process implemented during development which submitted daily production workload to MPI and generated comparative results for analysis.

Automated Analysis Data

```
hct391mcirrd001:~/mpi/daily_runs/20130810000000-
20130810235900$ ls
Algorithm20130810-20130810.csv
AlgorithmResults.20130810-20130810.csv
BaselineResults.20130810-20130810.csv
compare.20130810-20130810.xls
McirPd.20130810-20130810.log
McirPd.20130810-20130810.log.csv
McirPd.20130810-20130810.log.txt
McirPd_MpiDev.20130810-20130810.log
McirPd_MpiDev.20130810-20130810.log.csv
McirPd_MpiDev.20130810-20130810.log.txt
McirPD_MpiDev_compare.20130810-20130810.log
mpi_daily_job_20130816003002OO.log
SearchCriteria.20130810-20130810.csv
```

Automated Analysis Data

Search Fields	Queries	MCIR Match Found Count	% of Total Queries	MCIR Response (ms)	MPI Response (ms)
LegalLastNm,BirthDt	78,243	55,562	30.2%	46	227
LegalFirstNm,LegalLastNm,BirthDt	63,285	46,854	24.4%	45	489
LegalFirstNm,BirthDt	20,640	11,163	8.0%	47	325
McirId	20,594	19,547	7.9%	20	-
LegalFirstNm,LegalLastNm,BirthDt,SexCd	20,200	12,416	7.8%	47	468
LegalFirstNm*,LegalLastNm*,BirthDt	13,574	6,260	5.2%	73	462
LegalFirstNm,LegalLastNm	11,090	5,543	4.3%	82	399
PatientId	8,069	6,515	3.1%	349	-
LegalLastNm*,BirthDt	7,665	4,236	3.0%	70	272

Automated Analysis Data

20130910132204700 Reporting start time		
20130917114222400 Reporting end time		
Count	PCT	
244,892		Total searches
129,325	52.81%	Searches sent to MPI
70,112	54.21%	MPI search results used of those sent to MPI
70,112	28.63%	MPI search results used of total searches
811		Average search time (ms) where MPI search results used
438		Average search time (ms) where MCIR search results used
83		processed by PatientService while PatientService
358		Fastest search (ms) where MPI search results used
6,588		Slowest search (ms) where MPI search results used
320	0.13%	suspended
705	0.29%	suspended
78,183	31.93%	PatientService suspended
9,752	3.98%	Possible candidates returned from MPI where MCIR found nothing
219,750	89.73%	Searches with Fetch Size 10
25,142	10.27%	Searches with Fetch Size 200
3,000		MPI timeout (ms)
5,000		PS timeout (ms)
62		AlgorithmId

Logs from June 17-19, 2013
Raw Data:

- 104,911 searches
  - 141,709 MCIR results
  - 492,783 MPI results

Summary information:

- ~ 4:1 ratio of MPI search results to MCIR search results
- Almost 80% of highest MPI score per search had a MCIR match
  - When there wasn't a match: 63% used only 2 search parameters
  - Based on spot-checking the MCIR IDs, many of the non-matches were the same person with duplicate MCIR records.
- There are spikes in matching results at the following MPI scores: 65, 85, 91, 108, 114 (see graph below)
- 0.7% of search results with scores less than 65 had a matching MCIR search result.
  - 11.2% of searches with a score of 65 had a matching MCIR results.
- 21.6% of search results with scores ranging from 66-84 had a matching MCIR result.
- 74.6% of search results with scores ranging from 85-90 had a matching MCIR result.
- 93.6% of search results with scores over 90 had a matching MCIR result.

Describing the Searches (n=24,833)

- 13,877 (12.2%) of the searches did not have search parameters that meet the MPI criteria; these were deleted. ("Fault: no buckets were generated from search input data. nothing to search."
- 25,856 (24.6%) "duplicate" searches deleted
  - "duplicate" based on firstname, lastname and DOB variables
- 7511 (9.5%) had a wildcard in the first name
- 7749 (9.8%) had a wildcard in the last name
  - 4684 had a wildcard in both first and last name
- First name mean length: 5.89 (range: 1,8) (removing those with wildcards or no first name entered.)
- Last name mean length 6.23 (range: 1,8) (removing those with wildcards or no last name entered.)

MPI Results (n=492,783) Number of matches per search:

mpi_match	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	86354	17.52	86354	17.52
2	67958	13.79	154312	31.31
3	58841	11.94	213153	43.25
4	52429	10.64	265582	53.89
5	47108	9.56	312690	63.45
6	42634	8.65	355324	72.11
7	38874	7.89	394198	79.99
8	35654	7.24	429852	87.23
9	32737	6.64	462589	93.87
10	30194	6.13	492783	100.00

MCIR MPI – Next and Future Phases

- Currently in planning phase for submitting real-time adds/updates of MCIR person information to the MPI.
- Effective HL7 interaction with the MPI requires realtime interface to the MPI for adds/updates.
- Implementation of MPI for SOM HIE provides ability to add MPI search result to the HL7 message request as it is routed to the MCIR.
- Long term plans include using MPI to perform “member compare” during retrospective de-duplication / match-merge operations.
- SOM MPI may provide a low/no-cost address verification / cleansing service. Still TBD.

Community Toolsets

- Visit community web site for additional information / tools (temporary parking spot until coordination with related IIS programs takes place)

https://sites.google.com/site/iismatching

- Use of Google protocol buffer is an excellent solution for search logging where performance and storage considerations are paramount. Discuss with your technical teams.

MCIR – MPI Project Contacts

Gerry Bragg – MCIR Database Architect
Gerry.Bragg@Altarum.org
(734) 516-0825

Dennis Gilsdorf – MCIR MPI Architect
Dennis.J.Gilsdorf@Gmail.com

Bea Salada – MCIR Coordinator
Saladab@Michigan.gov
(517) 335-9340

Therese Hoyle – MCIR 'Consultant'
Therese.Hoyle@Gmail.com