Dynamic heterogeneous ecosystems

Semantic Systems Group
Professor Markus Stumptner
Dr. Wolfgang Mayer
Dr. Georg Grossmann
Dr. Jan Stanek

University Of South Australia

The presentation will have 2 parts:
In the first part I will give an overview of some projects we have done
In the second one I will draw a concept of collaboration in terms described by ecosystems paradigm.

Introduce who we are;

Markus – Leader of the Group and director of Advanced Computing Research Centre with years of experience in AI, constraint base modelling, ontology reconciliation and semantic interoperability. Multiple project under his lead were collaboration with major industry partners (Siemens, ANSTO, AutoCRC, CIEAM, DSTC) and had practical impact.

Wolfgang – is interested in the application of artificial intelligence techniques to solving practically relevant problems. Wolfgang made significant contributions in the areas of data modelling and knowledge representation methods, software systems interoperability, automated fault isolation in computer programs, and methods for semi-automated product and process configuration.

Georg – is working on the integration of business processes and complex data structures for systems interoperability and has applied his knowledge successfully in industry projects in collaboration with ANSTO, the Oil & Gas Interoperability pilot and local Adelaide companies such Mainpac and AOM. Recent work focusses on dynamic processes and complex event processing in simplifying the interoperability of ecosystems.

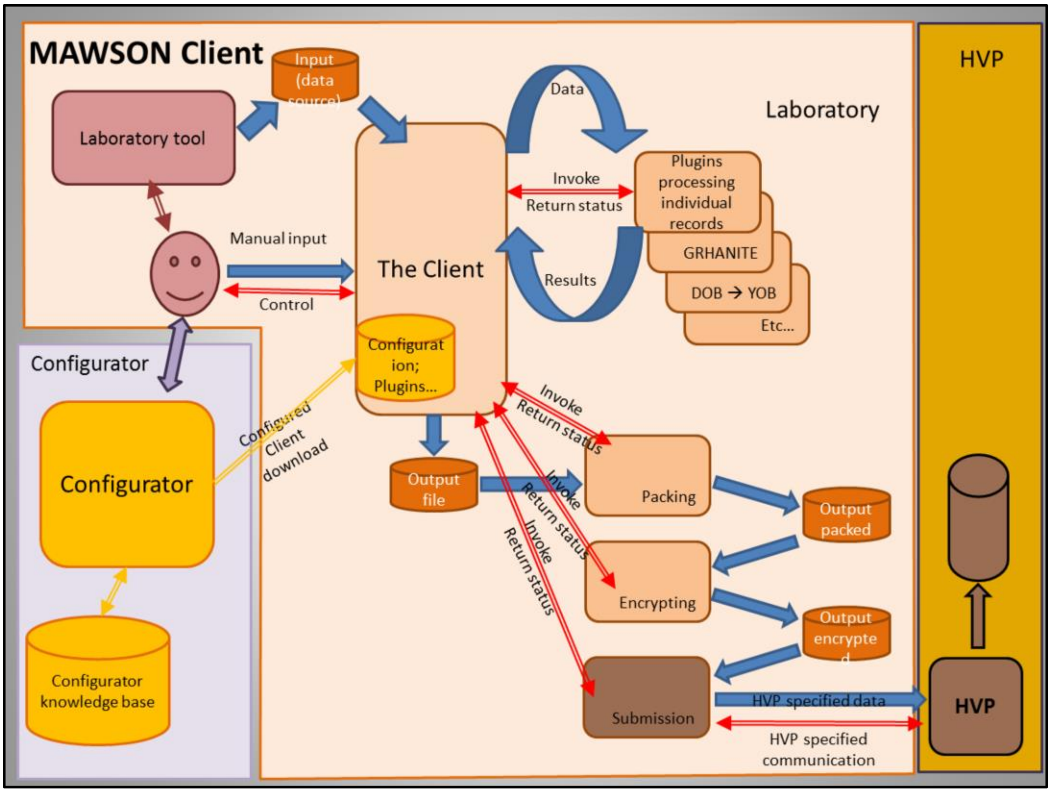
Jan – has major medical background (senior physician and consultant in internal medicine, head of department of clinical nutrition) and IT (research in AI, data analytics, clinical trial design and monitoring etc.) interested in clinical informatics and translation of informatics research results into clinical practice – in particular capture and analysis of data and processes in clinical settings and their utilization in care.

What we have done and do

Recent projects in health informatics

	Mawson Client	Prescription pathways	ED process mining
Area	Genetic testing	General practice	Hospital - ED
Purpose	Flexible data capture; Lightweight system integration	Visualisation of prescription patterns; Anomaly detection	Visualisation of patient flow through ED; Analysis of results
Stage	Tested prototype	Proof of concept	Report
Platform	Java, JavaScript	Cache/Ensemble	ProM, Disco,

This table shows a brief overview of projects I want to mention here. As you can see, they were positioned in different areas, but all of them were a reaction to a specific need:
Mawson Client – for automation of data collection to submit genetic testing results to a database
Prescription pathways – to extract additional value from prescription data ion general practice – at the practice level as well as for individual patient care
Emergency department processes mining and visualisation – visual representation of the processes in ED, and basis for further analysis of these flows.



First project I want to show is the Mawson client

The main driver for this tool was to get data from genetic laboratories safely, efficiently and without unnecessary exposure of confidential information. As no two laboratories had data in the same form, the tool needed to be highly configurable. Several versions of the tool were developed and successfully tested (thanks to QUPP, ANDS and NeCTAR funding

The current version has following properties:

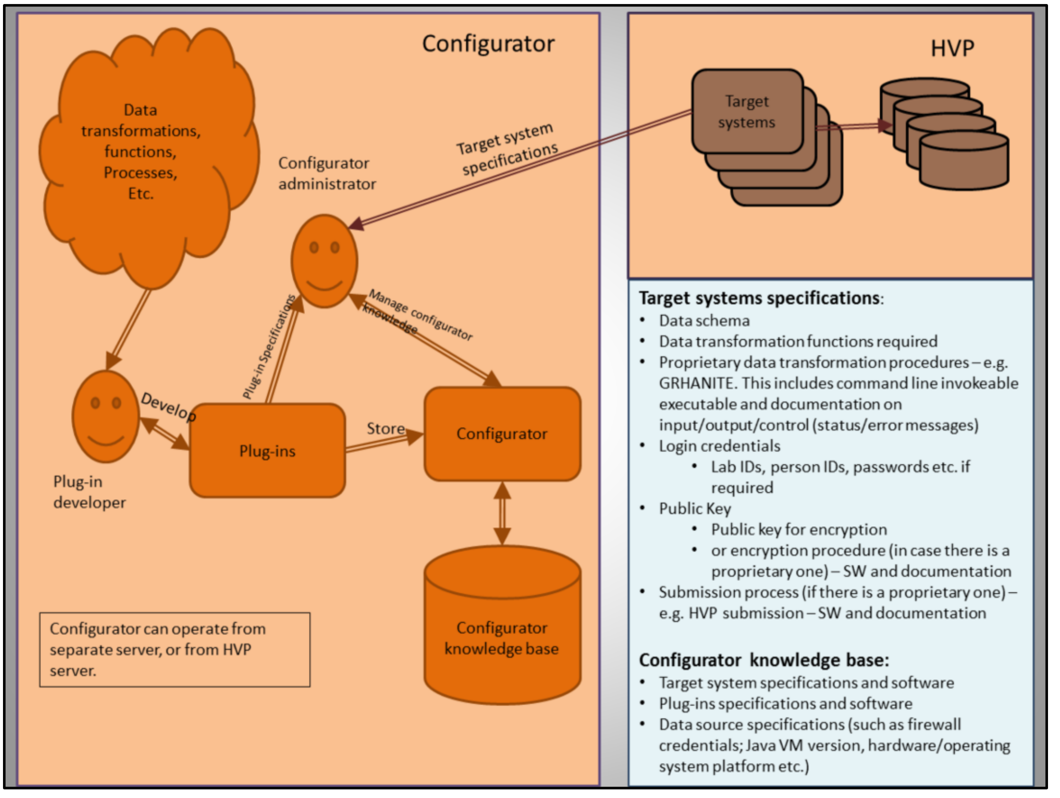
- Runs on client's computer without changing the computer setup (no registry entries etc.). Standard Windows XP machine is sufficient.
- Once invoked by saving a file into monitored directory, it will
    - read the file,
    - perform necessary data transformations (including actions to de-identify the data);
    - discard any data which is not required,
    - collect necessary additional data manually;
    - pack and encrypt the resulting dataset,
    - log onto the target system and
    - submit the dataset.

Why I am showing this – why is this is relevant?

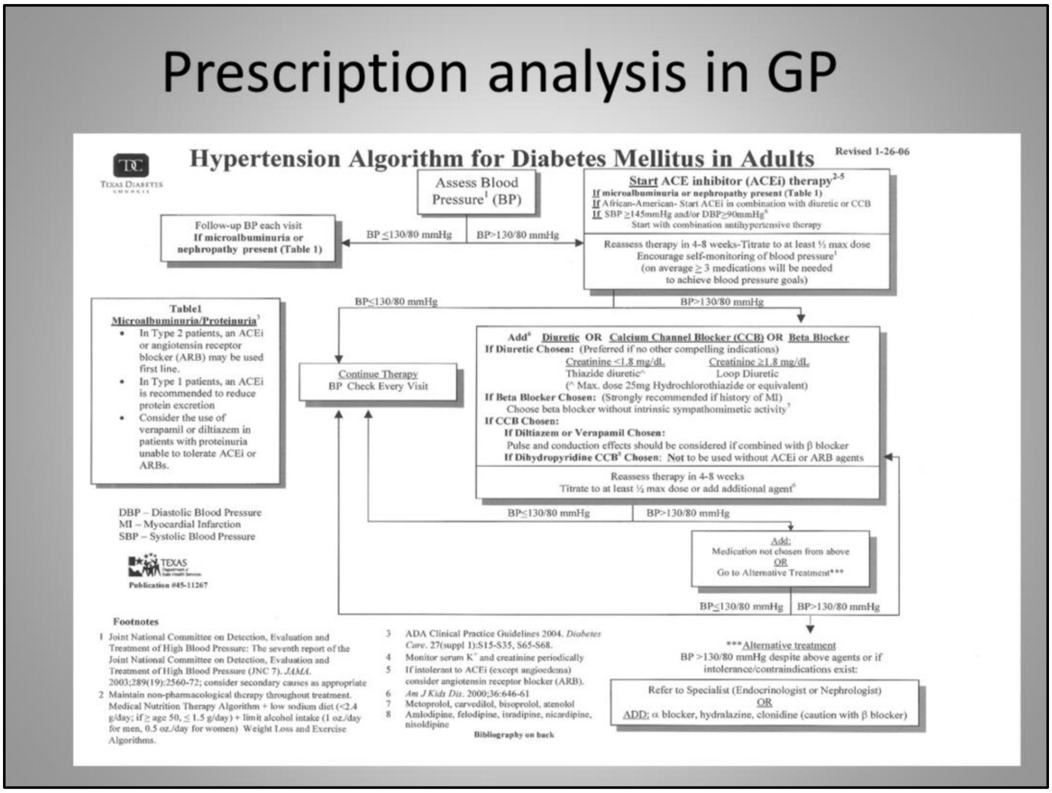
The tool provides lightweight integration to connect heterogeneous systems which are not capable of communicating directly. At the same time the tool performs data schema reconciliation between the sending and receiving system, allowing to collect missing data manually.

What do we solve with this?

1. Recover control flow from data flow
2. Protect privacy (sanitization runs on client's computer)
3. Safety (encryption does not rely on VPN only; log-in onto server)
4. Manual data entry optimized
5. Low cost
6. Self-configuring (does not require IT support to configure the data collection tool – configurator)



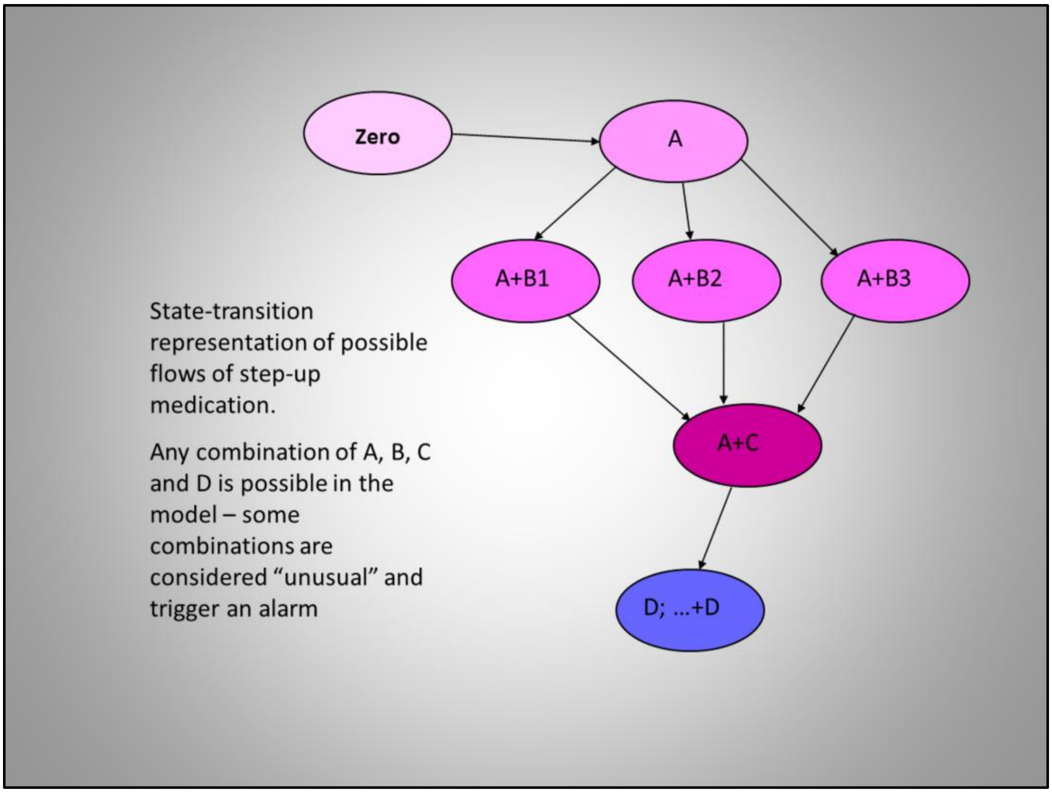
The most recent contribution to this tool is the configurator. This component will take a sample dataset and guide the user through a series of steps to configure reconciliation of data schemas between the source and target systems, set up data transformations (simple ones can be set up by writing a chunk of Java Script code; repetitive transformations or more complex ones can be included as plugins) and define credentials for firewalls and target server. Upload to the target server is done via submission plugin. (e.g. for data submission to HVP we used a submission process comprising GRHANITE proprietary algorithm. This process was implemented as a specific plugin written by colleagues with access to the proprietary source code)



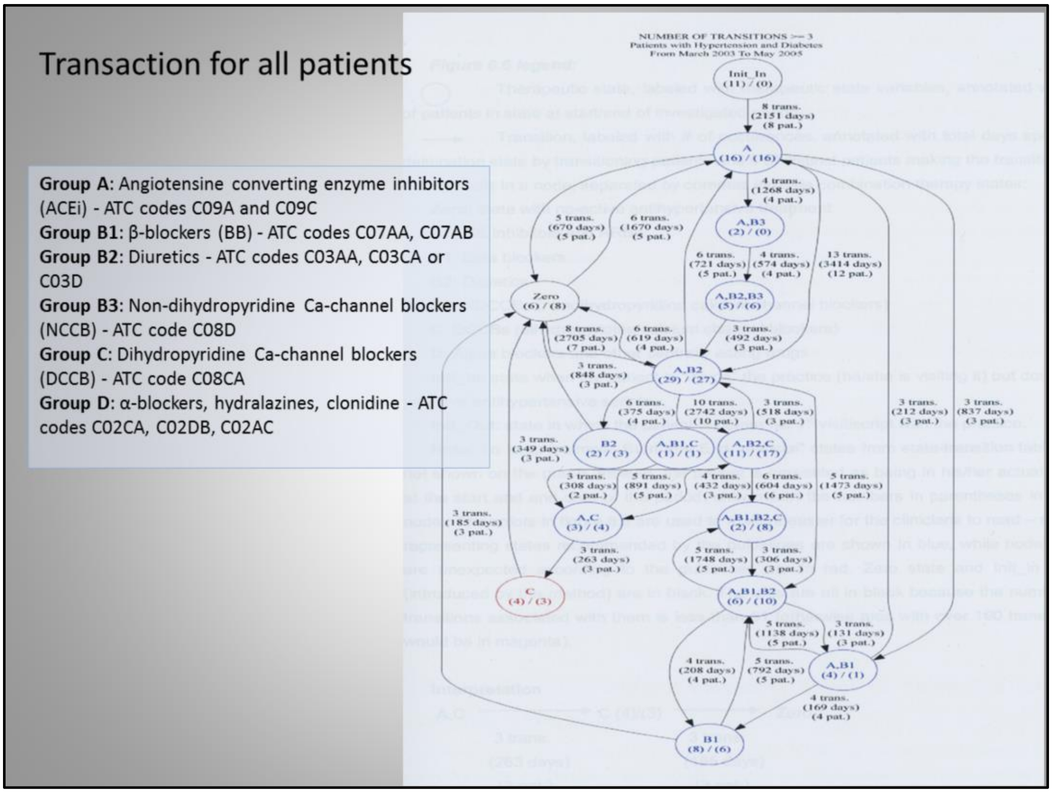
Another project I want to mention is modelling and analysis of practice processes

The idea behind the project was straightforward: Prescription reflects significant part of clinical practice (and prescription data is side product of practicing medicine, the quality of data is typically good and the data itself is computer readable).

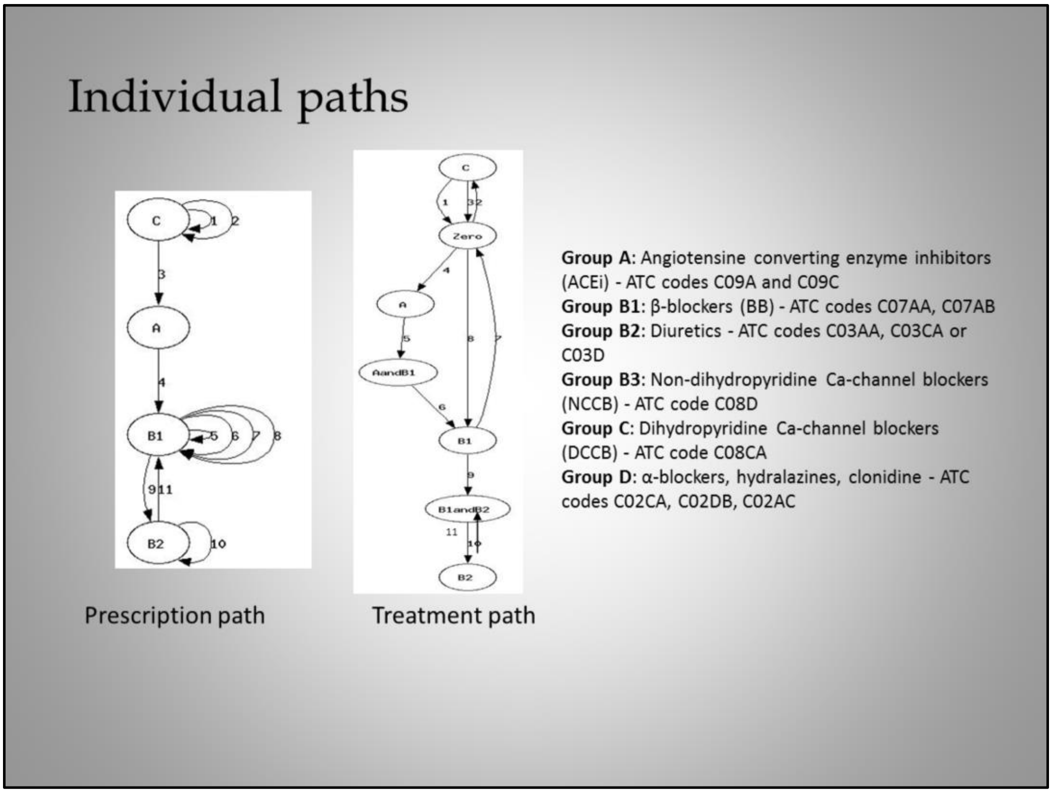
To restrict the complexity – the project was set up as proof of concept – we used a simple guideline as a basis for definition of states and actions.



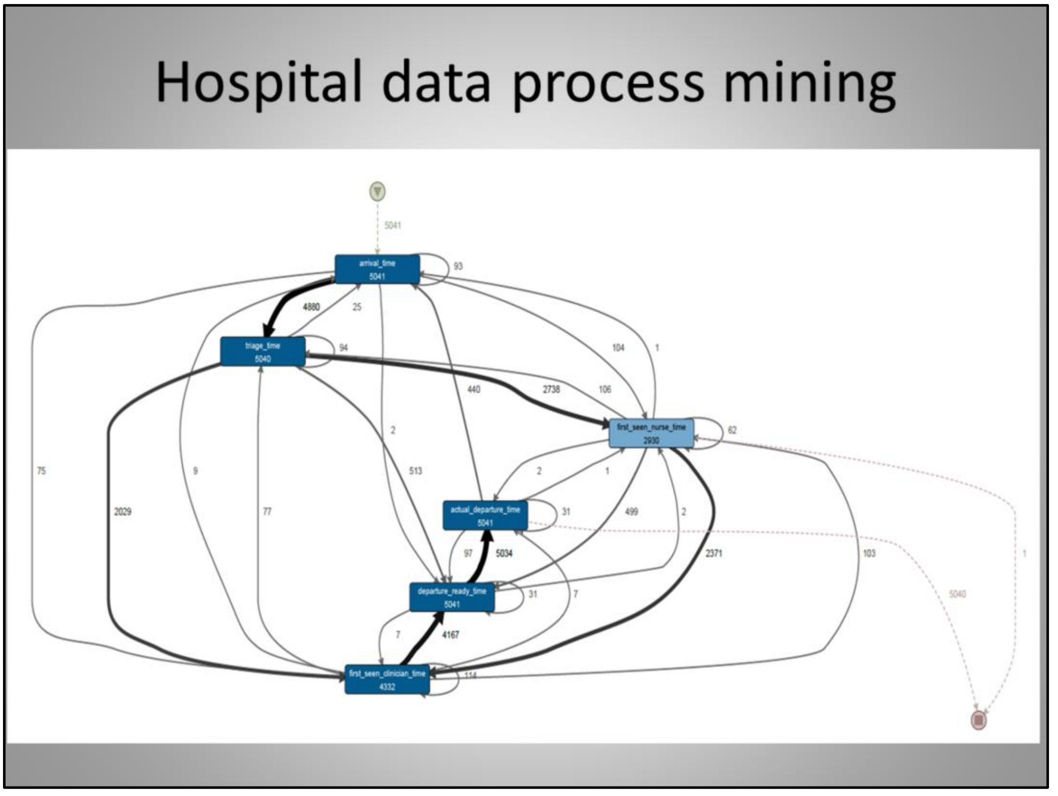
The guideline was represented as a state-transition graph – this slide shows the states as previewed by the guideline. Obviously – several other states are possible…



Once the actual data was fitted into the model – letters A, B C and D stand for categories of treatment – we obtained a graph representing prescription and its changes across all patients in the practice.
Analysis of such graphs from the collaborating practice showed that vast majority of nodes conformed to what was expected by the guideline; the main course of treatment was following the pattern recommended by the guideline. However, as can be seen in this picture, we could identify aberrant nodes – such as the red one showing monotherapy with Dihydropyridine Calcium channel blockers.
This information was then used by the collaborating general practice to focus patient record reviews on the unusual or aberrant cases, as well as assess the overall conformance with the guideline (i.e. use of the information to improve individual care quality, as well as overall practice quality and safety management).



Extension of the project was extraction and visualisation of individual patients – here we see comparison of the prescription pattern (i.e. what drugs were prescribed and when) and a treatment pattern (i.e. what we assume the patient was actually taking).
As can be seen from this example, from looking at the prescription in the patient record (or looking at the visualisation of this information) is not sufficient to identify possible problems. However, by looking at the treatment pattern we can readily see the patient was several times with no treatment (Zero state), as well as on different combinations (A and B1; B1 and B2). Extracting these patterns, presenting them in visual form as well as using them as a basis for targeted alert generation was shown in clinical trial as a useful contribution to quality of care and patient safety.



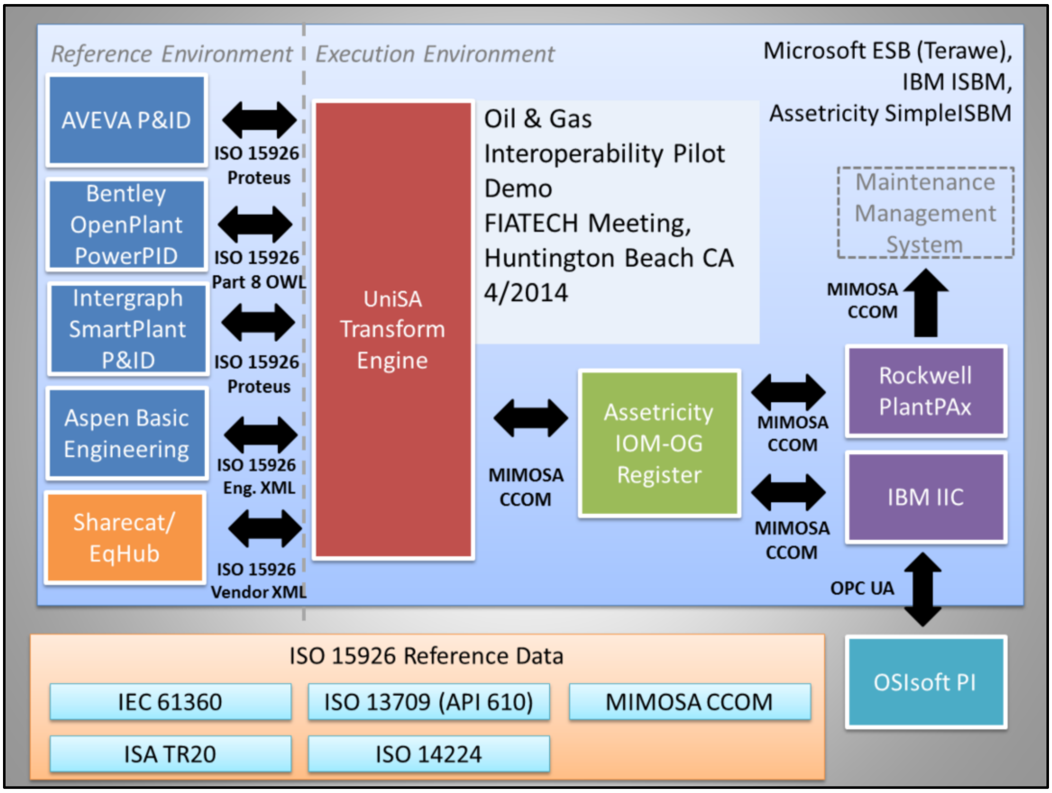
This project is an example of process mining

As an example of process mining I would like to show a visualisation of processes at a hospital emergency department. The assumed process is simple: Patient arrives at the ED, is triaged, seen by nurse and/or clinician, problem resolved – ready for departure and finally patient leaves the ED.

And indeed, - as can be seen by the thick lines, the real process unfolds along the expected path.

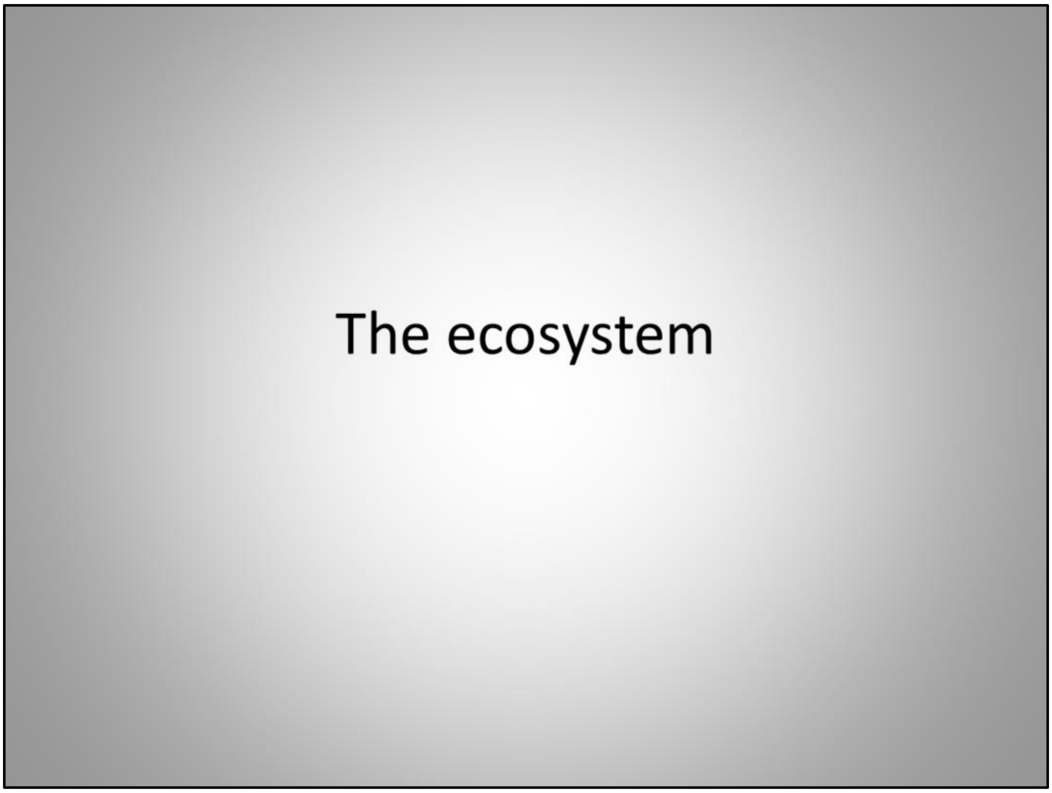
However, there is a whole cornucopia of alternate links, some of them evidently unusual. This result can be used just as a simple visualisation of what happens at the ED (for managers to ponder), but it can be also used as basis for analysis such as:

- Data error detection (e.g. loop to the same node within seconds is coding error – duplicate event entry)
- Unusual pathways – patients abandoning the process, but returning at a later stage (link from a non-terminal node back to “triage” or to “arrival”…)
- Distillation and further analysis of subset of patients travelling along unusual or undesirable pathways (can have implications on patient safety and quality of care … - e.g. analysis of pathways of patients returning to ED multiple time).
- Analysis of resource utilisation (especially if localisation is added) – beyond simple counting of cases using a particular service, but offering more context and a more dynamic view on usage of a particular service ore set of services…)



The projects I showed so far have their challenges but were comparatively small. At this stage we do not have any large and complex project in health interoperability - “courtesy” of health data unavailability (and this is indeed one of the reasons why we are here…). Instead I want to show a large and complex problem developed in close collaboration of Semantic Systems group with several major partners from oil and gas industry.

Semantic systems group developed a prototype of an integration engine allowing integration across organisations using different standards, different processes, different terminologies within the given domain. This solution spans across multiple organisations, multiple standards and a wide diversity of objects (plants, rigs, etc.) and I will return to this in more detail.



Once the complexity of the information system exceeds certain level, we may consider to change the paradigm. So let’s return back to health.

Providing health care, maintain wellbeing

Service needs to be:
- Safe
- Appropriate
  - Up-to-date
  - To high standards
  - Personalised
  - Holistic
- Continuous
- Overarching
- Transparent

What do we expect from health care?

In the first place Clinical practice should be safe - “noli nocere” is the ancient paradigm of medicine.
It has to be appropriate – i.e. provided at the current level of knowledge, to current standards. However, this is not sufficient – we need to look beyond the general guidelines and standards and appreciate the personality of the patient – in her own historical, cultural, family etc. context.
There is a lot of talk now about patient journey – as a metaphor – the health service starts at the time an individual is conceived, and ends after the person has died – this time now spans over 80 years (and growing).
Service provided should go beyond treating disease – the overarching concept is to cover factors influencing out wellbeing before this leads to recognisable disease. Indeed, this includes – by direct or indirect influence – all areas of human activities and beyond…
Last but not least – the service needs to be transparent, so that it can be understood by all stakeholders involved – especially our main partners in fighting disease and health problems: -- the patients.

This list is arbitrary and does not cover all details, but should be enough to demonstrate the width of scope we face when dealing with health and wellbeing services.

On the top of this – many fields evolve, emerge and disappear at an increasing pace, so the ecosystem we look at is highly dynamic. Genetic testing – as an example – was a bit outlandish, highly specialised research field; now it is rapidly becoming a cornerstone of personalized medicine – clinicians, geneticists, bio-informaticians and theoretical mathematicians have to work together in unprecedented coalitions and we struggle with ethical implications of the possibilities these methods offer.

Goals of health care
http://www.safetyandquality.gov.au/wp-content/uploads/2012/08/Goals-Overview.pdf
http://www.searo.who.int/entity/health_situation_trends/documents/MDG_Brochure_2012.pdf?ua=1
http://www.ncbi.nlm.nih.gov/pubmed/10186481 **Changes in the healthcare system. Goals, forces, solutions.**

How we see the status quo

Current situation is characterised by:

- Governance/funding splits (commonwealth, states)
- Institutions (universities, hospitals, GPs, pharmacies, aged care providers, community services etc…)
- Diversity of specialisations in health
- Ownership (public/private/hybrid)
- Geography (urban/rural)

Challenge proclaimed: break out of the silos!

Looking at the health care system– it is strongly aligned around historic silos – some of them are specific for Australia, some are not; many of them exist for a good reason.
These were created along the lines of governance and funding:
states/commonwealth (splits the healthcare system horizontally – hospitals are managed by states; GPs by commonwealth etc.);
Individual institutions (with their own internal systems driven by the institution’s goals and objectives, culture and traditions etc.)
Specialisations with information systems tuned to suit their specific needs
Ownership models: private/public;
Geography: urban/rural

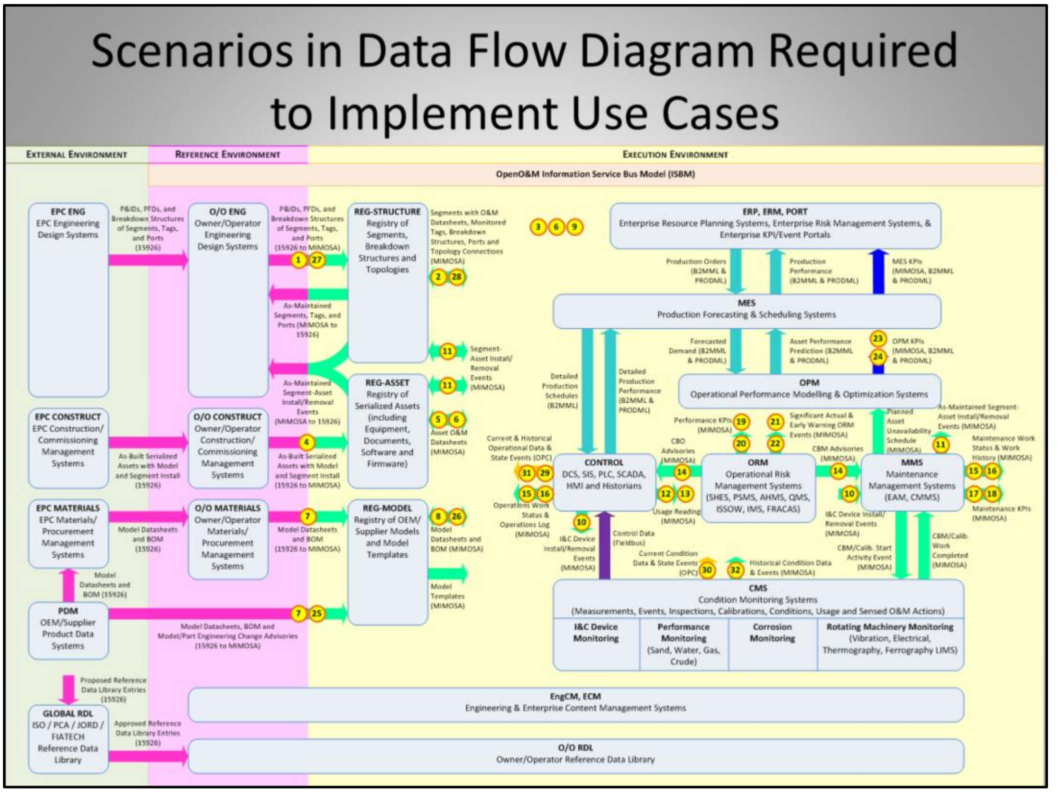
Breaking these silos and offering integrated care is one of the major proclaimed goals (and challenges) of PHNs – BUT – many of these divisions are meaningful and removing them will be counterproductive.

Ecosystem	Engineering
Decentralised control	All conflicts must be resolved (and resolved centrally and uniformly)
Inherently conflicting, unknowable and diverse requirements	Requirements can be known in advance and change slowly
Continuous evolution and deployment	System improvements are introduced at discrete intervals
Heterogeneous, inconsistent and changing elements	Effect of change can be predicted sufficiently well Configuration information is accurate and can be tightly controlled Components and users are fairly homogenous
Erosion of the people/system boundary	People are just users of the system Collective behaviour of people is not of interest Social interactions are not relevant
Failures are normal	Failures will occur infrequently and defects can be removed

As many of the silos make sense and should not be removed, a different approach needs to be considered.
This table shows the characteristics of an ecosystem coming from the Carnegie Mellon Information Systems Institute:

I think we can agree, that the characteristics on the left are a good fit to characterize the situation we have in health care:

- Decentralization [multitude of systems]
- Inherently inhomogeneous, even conflicting goals and requirements (e.g. data mining for knowledge generation – or in extreme case – data analytics to improve safety vs. privacy)
- Continuous evolution (and deployment) – in fact the classical cycle of distinct phases might disappear (individual components evolve at their own pace)
- Heterogeneous, inconsistent and ever-changing elements (e.g. HL7 v.2 vs. v.3 vs. FHIR…)
- Erosion of the people/system boundary (e.g. augmented reality, wearable computing, Internet of Things…)
- And failures of components – like it or not – will be part of normal life of the systems given their growing complexity and pervasivity



Going back to the oil and gas industry example:

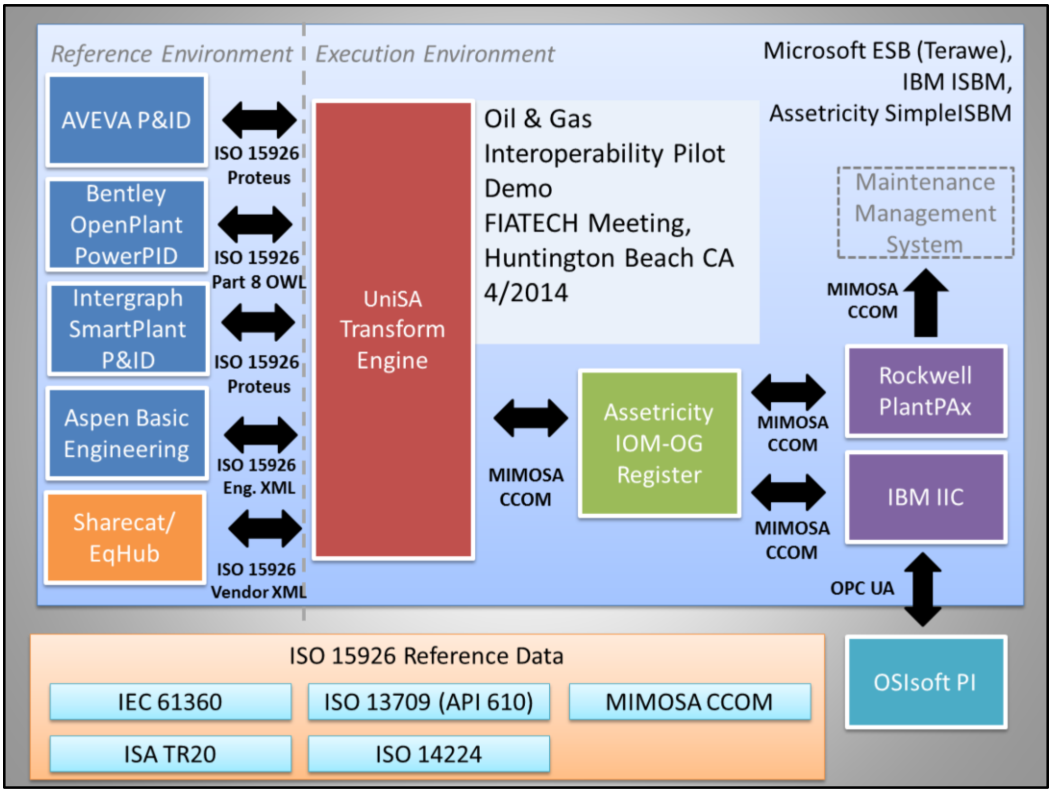
We can see a diverse set of functional components owned and controlled by disjunct organisations required to work together on projects – possibly creating ad-hoc alliances to achieve specific goals. Each of these companies is using more-or-less mature systems which reflect specific needs of the particular organisation and utilises a particular set of standards – both systems and standards evolve, change in time (so there is no guarantee that the components shown here will be using the same version of the standard).

These systems cover information from planning and design, through to operations and maintenance. Typical flow of information is from design to operations to maintenance. The transfer of information is manual, or semi-automatic; documents are produced, printed, circulated, read, re-typed etc. The process includes a great deal of human involvement.

Typical direction of information flow is one-way -- any change done downstream is seldom backfilled (e.g. a change on a plant done at the time of maintenance may not be reflected in the design documentation unless there is an explicit requirement to do so). The result is divergence between the documentation in different categories (design, operation, maintenance etc.) even if they use the same standard (which they don’t).

At the same time, constraints under which operations unfold, are dictated by external environment (such as legislation or regulatory framework), and evolve. These constraints need to be taken into account as well.

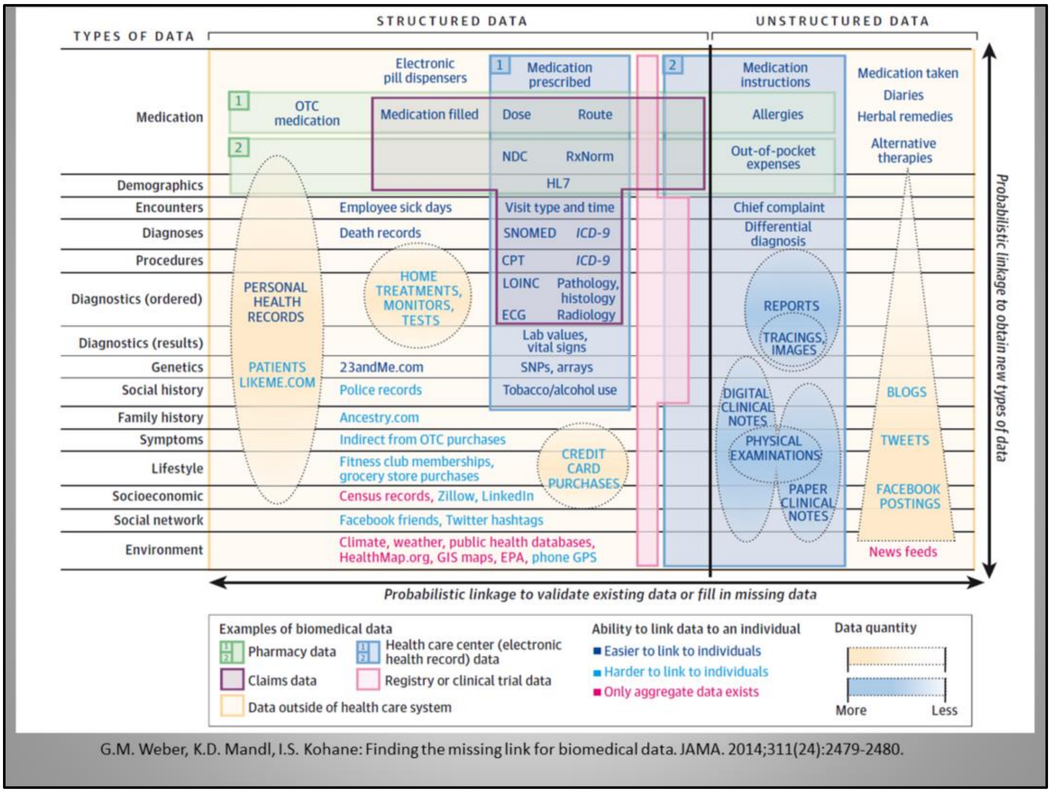
The result is a highly heterogeneous system – heterogeneous both in terms of structure and contents, which is operating under a diverse and changing set of constraints. In such situations the traditional engineering paradigm may not be optimal.



Ecosystem in practice – existing mature systems needed to be integrated:
Integration in both direction – from design through production to maintenance, but also back from maintenance to production and design (to keep the documentation, policies, standards etc. updated all the time).
Integration across several standards – where *meaning* has to be preserved.

To achieve these goals we created mechanisms for abstract description of ontologies and their transformations; as well as description of processes, process characteristics and behaviour. Finally we assembled these components into a hierarchy of models. This approach allows higher level information (data, processes) ecosystems analysis and integration. Formal methods can be used to support consistency and integrity of the resulting integration engine further reducing the manual maintenance effort required.

Full discussion on the pilot demo:
https://www.youtube.com/watch?v=ShvX4_C7QJg (WARNING: for dedicated engineers only! ;-))



[slide - G.M. Weber, K.D. Mandl, I.S. Kohane: Finding the missing link for biomedical data. JAMA. 2014;311(24):2479-2480. doi:10.1001/jama.2014.4228]

What is the situation in heath? The current rage in the United States is Health Information Exchange – it is no more sufficient to manage electronic health records within one organisation. The requirement is to share relevant pieces of information with others caring for the same patient and with the patients themselves.

The wealth of possible information sources is summarised in the table recently published in JAMA. I would like to point out the “non-medical” data sources which were included into the table – while not being created and used for health use (and hence not subject of any specific regulation or standards) they can still contain useful information. Good example of this can be as adverse drug reactions extracted from Tweets…

Some of these systems are solid, stable, well designed, but many are volatile and rapidly developing with multitude of emerging – not always predictable – interactions (as an example – health apps for mobile phones predate long term monitoring for diagnostic purposes; yet are rather ephemeral and subject to fashion).
There is no governing body operating across all of these components, nor a standard, legislation, policy etc.

As a result, we perceive the need to see this as a highly dynamic, heterogeneous complex ecosystem of services, systems, processes, data and metadata.

So the challenge is not just to break, or better to say - bridge the silos. The challenge is to understand this ecosystem, so that we know what to do to make it stable and healthy.

Our team wants to be part of this process and bring to table our capabilities

- in ontology analysis, reconciliation;
- in process extraction, analytics and modelling;
- in advanced analytics of data (including big data – UniSA is part of the Data to decisions CRC) and, ultimately,
- complex system modelling and integration.

As such, we look for long-term relationships with industry partners, so if you think you have problems we can share or help you with, we will be pleased to hear from you.