



Съвременни пресмятания за иновации  
AComIn: Advanced Computing for Innovation

# Автоматично извличане на пациентно ориентирани данни от свободен медицински текст на информатизирани здравни записи

**Доц. д-р Димитър Чаръкчиев, дм**

УСБАЛЕ “Акад. Ив. Пенчев”, Медицински университет – София

**Проф. Галя Ангелова, дмн**

Институт по информационни и комуникационни технологии, Българска  
академия на науките



*AComIn Workshop*  
*15-17 May 2013*

# БОЛНИЧНАТА ИНФОРМАЦИОННА СИСТЕМА НА УСБАЛЕ

- Подобряване на лечебно-диагностичната и профилактична дейност и
- Постигане на ефективно управление, **посредством**
- Усъвършенстване на информационните процеси чрез автоматизирането им

# ОБХВАТ НА БИС

- Регистратура и каса
- Приемни и консултативни кабинети
- Клиники и отделения
- Лаборатории
- Образна диагностика
- Болнична аптека
- Администрация
- Връзка с МЗ, НЗОК, РЦЗ и др.

# ИНФОРМАТИЗИРАНО ЗДРАВНО ДОСИЕ

## Досието на пациента е:

- **Точен и пълен структуриран запис на всичко** което е наблюдавано, обсъждано и извършено от лекаря ,
- който **осигурява обмена и повторното използване** на данните и знанията,
- на приемлива цена
- **Запис, в който са съхранени всички необходими сведения за лечението и наблюдението** на пациента
- **Физически разпределено и поделено** между здравните професионалисти.
- **Средство за комуникация и за управление** на свързаните с него процеси

# ИНФОРМАТИЗИРАНО ЗДРАВНО ДОСИЕ

## Електронен здравен запис (ЕЗЗ)

- Хранилище на личните здравни данни във формат позволяващ обработката им от компютърни системи
- Изградено на базата на модел на информацията (*ISO*)
- Автори на моделите на понятията в конкретните области са клиницистите
- Информатиците – автори на моделите на информацията

Основни данни на пациента

Имена на пациента: **Христо Антонов Димитров**      Възраст : **18 год**    Дата на приемане : **21.05.2008 г.**  
 Приет по клинична пътека : **Заболяване на хипоф.и надбъбр.под 18год**      Приемаш лекар : **Д-р Нато Василевски**  
 Приет с диагноза : **Кръвно налягане и други данни за пациента**  
 Лекуващ лекар :  
 Алергичен към :

**Кръвно налягане и други данни за пациента**

Ръст (в сантиметри) :     Тегло (в килограми) :     ИТМ :

Систолично :     Диастолично :     **Покажи допълнителна информация**

Размер на маншета :

Използван уред :

Място на измерване :

Метод на измерване :

Събитие :

Статус при измерване		Информация	
Поза на измерване :	<input type="text" value="Standing"/>	Систолично :	<input type="text" value="120"/>
Състояние при измерване :	<input type="text" value="At rest"/>	Диастолично :	<input type="text" value="60"/>
Ниво на усилие :	<input type="text" value="0"/>	Средно артериално налягане :	<input type="text" value="0"/>
Наклон при измерването :	<input type="text" value="0"/>	Налягане на пулса :	<input type="text" value="0.0"/>
		Коментар :	<input type="text" value="Без ортостатизъм"/>

Основна диагноза и пр

Основна диагноза : AI



Придружаващи

Данни за изписване

- Изписан
- Преведен
- Починал

отказ от лечение от ЗЗОЛ        Лекар :

Пролежани леглодни :     Болничен лист No :     Брой дни :     Състояние :

Изпиши пациента от лечебното заведение

13606-1

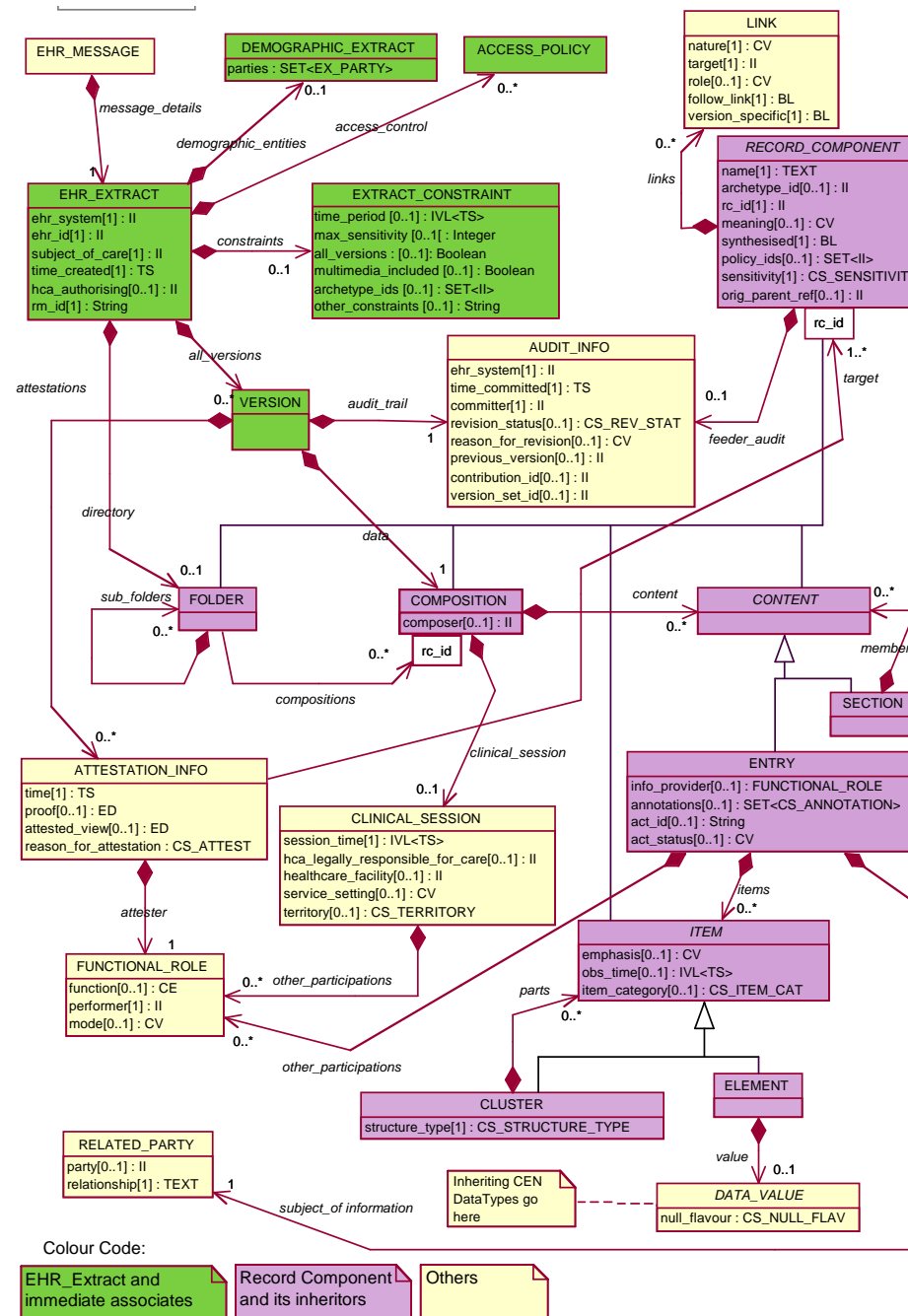
Референтен модел

**L'EHR\_EXTRACT е коренът на извадката от E33**  
- представен на *зелен* фон.

Компонентите на записа са представени на *виолетов* фон

**Всички други класове са представени на бял фон**

Цветовете са без значение за моделирането



# Архитектура на ЕЗЗ

- Полезните сведения в ЕЗЗ, са **структуриране йерархично**.
  - Медицинските наблюдения, цели и изводи могат да имат проста или по-комплексна структура.
  - Тези елементи, по принцип са структурирани в параграфи, съдържащи се в « *документи* » като например бележките от консултациите, протоколите и епикризите.
  - Тези документи са *подредени в класьори* (папки)
  - *Множество класьори могат да се отнасят към един пациент* в рамките на една болница (например класьор за медицинските грижи (история на заболяването), класьор с рентгеновите снимки, класьор с хистологични резултати и др.)
- Референтният модел трябва да отразява **тази структура и тази йерархична организация**, за да бъде верен на оригиналния медицински контекст и да гарантира ,че смисълът се запазва при обмен на ЕЗЗ между различни системи



# Контекст на информацията

- Стандартът използва определението Първичен контекст на информацията (Original Information Context) за да представи идеята, че всеки компонент на E33 има първичен контекст
- Всеки компонент на записа (Record Component) е вложен в идентифициран архитектурен компонент
- Възможните контейнери са:
  - **EHR\_Extract** – кореновия клас (**Root Architectural Component**)
  - **4 типа на оригинални комплекси от компоненти (Original Component Complex (OCC))**
    - Папка (Класьор)
    - Композиция
    - Секция
    - Група (Cluster)

# Йерархични компоненти на ЕЗЗ

**ЕЗЗ (EHR)**

**Папки (Folders)**

**Композици**

**Секции**

**Входни данни (Entries)**

**Групи (Clusters)**

**Елементи**

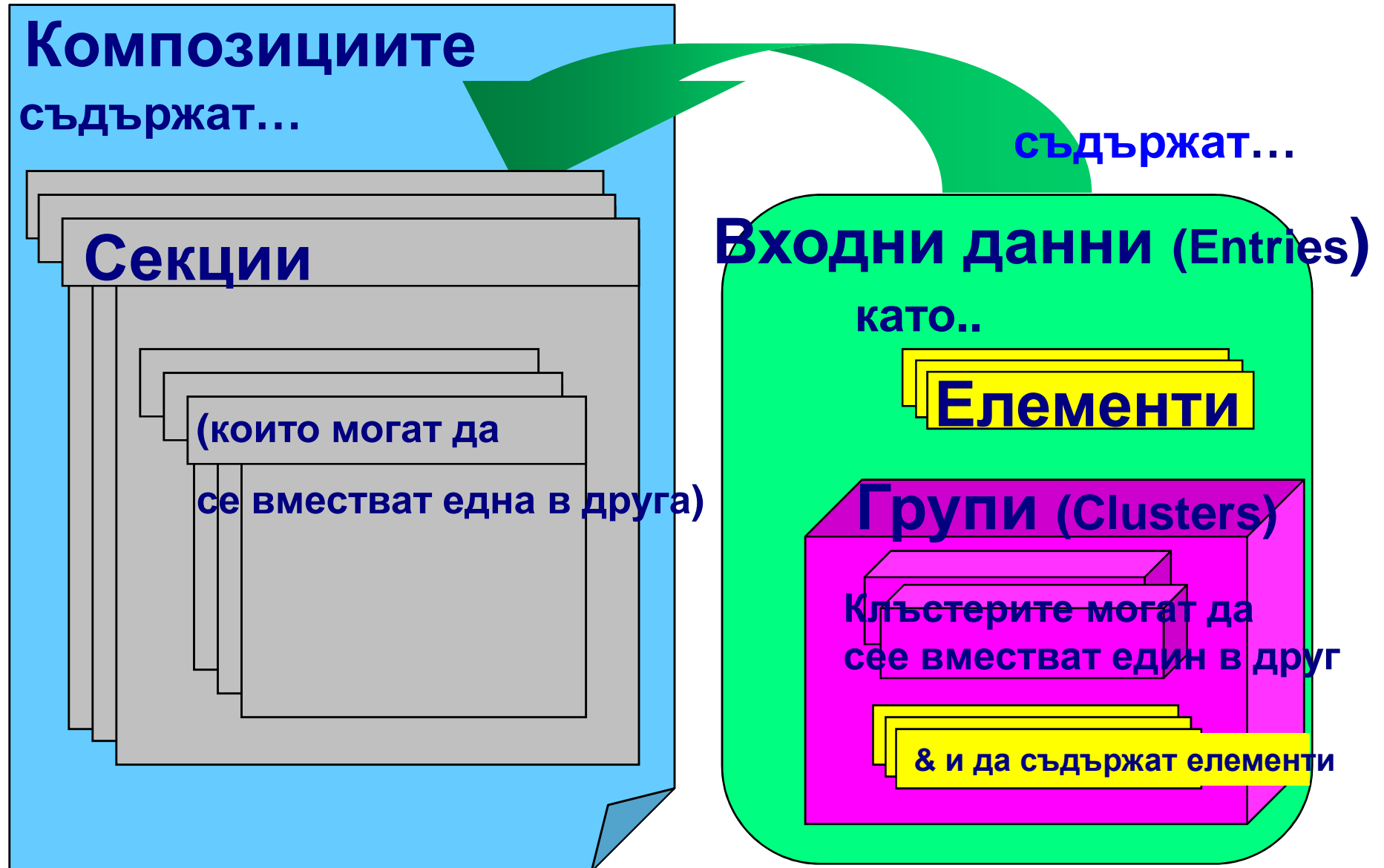
**Стойност на данните**

- Представява Електронния здравен Запис отнасящ се за един човек
- Най-високото ниво на ЕЗЗ, групиращо части от досието - по епизод, по мед. специалност и др.
- Клиничен преглед, посещение, или документ – напр. резултат, епикриза
- Параграфи представящи фазите или различни моменти на посещението
- Информация за наблюдение(я), оценка(и) и предписание(я)
- Комплексни структури (таблици, серии на времеви интервали), напр. интервали
- Крайно разклонение на йерархията, имащо определена стойност - пулс, тегло
- Тип на данните за всяка реална инстанция например кодирани термини , количества с техните единици

# Логически изграждащи блокове на ЕЗЗ



# Логически изграждащи блокове на E33



# Включени данни

- Данни за престоя – дати на приемане и изписване, продължителност, превеждане от една клиника в друга
- Данни за пациента – демографски данни, възраст, пол, състояние, ДСГ и др.
- Диагнози – МКБ 10
- Процедури – МКБ 9 КМ
- Медикаментозно лечение – АТС кодове
- Лабораторни изследвания
- Текстови файлове – епикризи

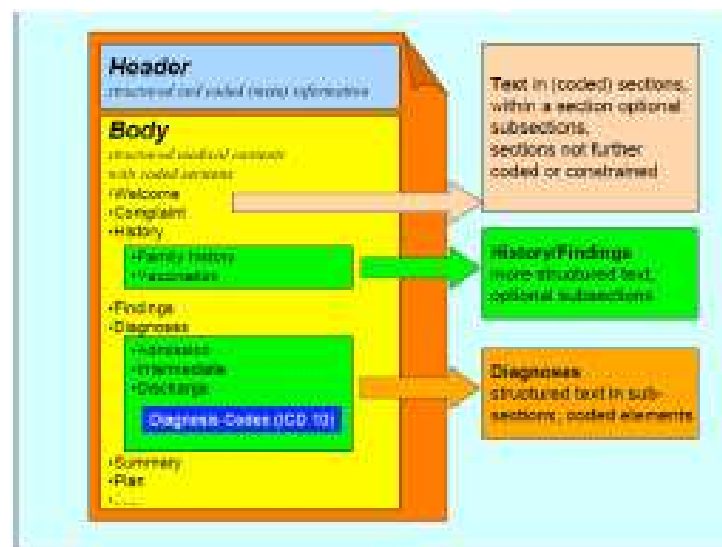
# Структура на епикризата

- Лични данни
- Диагнози
- Анамнеза
- Статус на пациента
- Лабораторни данни
- Консултации
- Обсъждане
- Лечение
- Препоръки

# Архитектура на клиничния документ

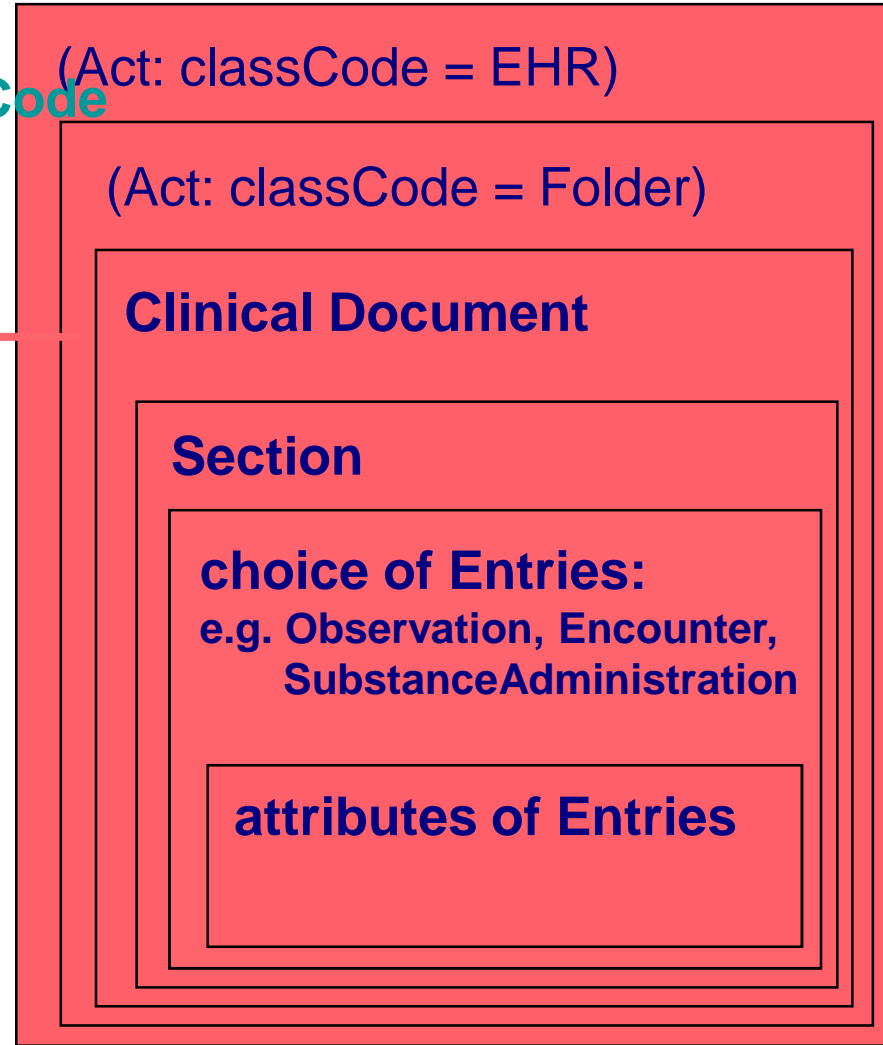
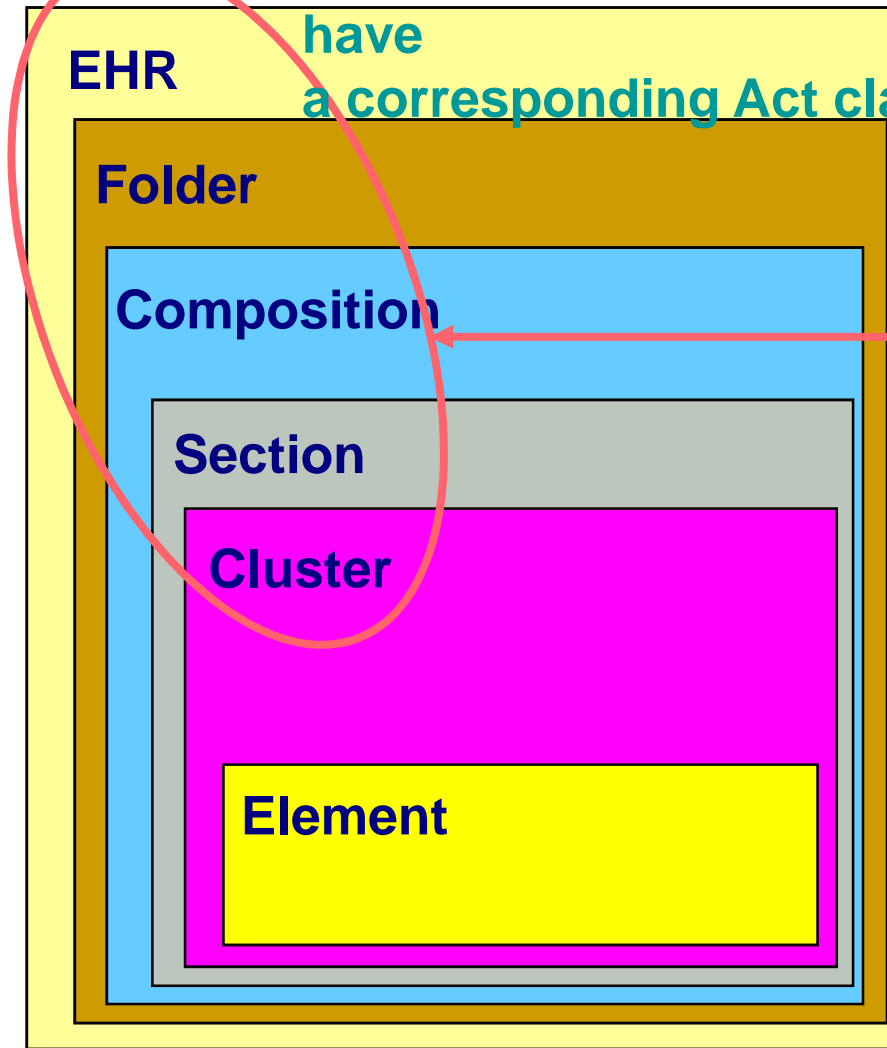
Clinical Document Architecture (HL7 CDA)

Прилага се в редица страни за форматиране на епикризи



# EN 13606 отнасяне към HL7 & CDA

All of the EN 13606 container classes now  
have  
a corresponding Act classCode





# Българският медицински език

- В здравните записи се използва медицинска терминология на български и латински език
- Няма предпочитания и едновременно като синоними се използват термини на двата езика
- Диабетес мелитус – типус 1. Полиневропатия диабетика. Angiosclerosis vas. retinae hypertonica. Retinopathia diabetica simplex.

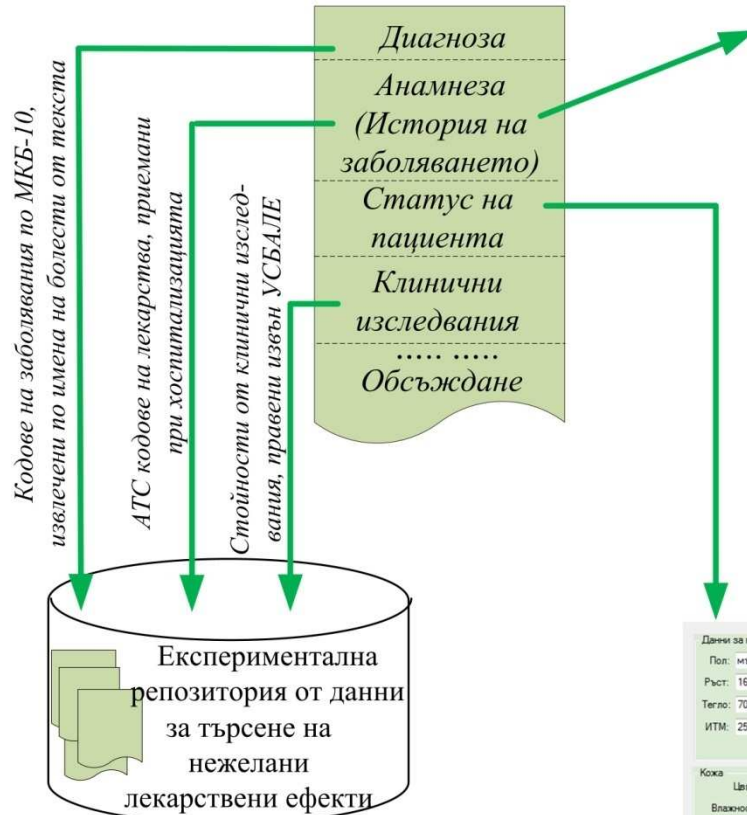
# Българският медицински език

- Понякога латинските термини се изписват на кирилица, особено когато лекарят няма желание да превключва клавиатурата от кирилица на латиница
- По принцип смесицата на латински и български термини е всеобщо приета, включително и в официални документи

# Автоматично извличане на данни от информатизираните болнични записи

- В рамките на два проекта “ЕВТИМА” и “PSIP” бяха разработени и тествани върху корпус от 6200 епикризи 3 компонента за автоматично извличане на медицински данни
  - *Компонент за автоматично присвояване на кодове по МКБ – 10 на диагнози*
  - *Компонент за автоматизирано извличане на данни за лечението, базиран на АТС кодове*
  - *Компонент за извличане на стойностите на лабораторни изследвания*

# 6200 анонимизирани епикризи



### Temporal Events Diagram

episode 1  
 episode 2  
 episode 3  
 episode 5  
 episode 6  
 episode 11

diagnosis  
 диабетна полиневропатия : G63.2 , Диабетна полиневропатия (E10-E14) с об

**Record ID: 100046-08**

**Данни за пациента**  
 Пол: мъж    Общо състояние: добро  
 Ръст: 165    Позиция в леглото: активна  
 Тегло: 70    Фебрилност: афебрилен  
 ИТМ: 25    Възраст: отговаря на действителната

**Двигателна система**  
 правилно развита

**Дишателна система**  
 Гръден кош: емфизематозен  
 Перкуссионен тон: хиперсонорен  
 Респираторна подвижност: запазена  
 Дишане: чисто везикулярно

**Сърдечно-съдова система**  
 Пулс: 78 уд./мин.  
 КН: RR 120/80 mmHg  
 Сърдечна дейност: ритмична  
 Тонове: ясни  
 Шум: систоличен на върха

**Отделителна система**  
 Succusio Renalis  
 Ляво: -  
 Дясно: -

**Долни крайници**  
 Откоп: пиповат  
 Пулсации на периферни артерии: запазени

**Общи**  
 Кожа: Влажност: много суха  
 Стилпала: фикуси

### Описание на лекарственото

Измисляне на медикамент    Редактиране схема    Прехвърляне на медикамента    Изтриване лекарството

1 пъти на ден в (08:00) по (10) IU.

Дата(Н)	Брой Дни	Дата(К)	Наименование	Кол. от Аптека	Кол. от Клиника	Кол. от Пациент
28/04/2011	7	04/05/2011	Инсулаглар НМ 4 IU 10Ммл	0.000	0.000	1.000
28/04/2011	7	04/05/2011	Рокатрол капсул. 0.25mgx100 бр.	0.000	0.000	7.000
28/04/2011	7	04/05/2011	Беталок табл. 50 мг	0.000	0.000	12.000
28/04/2011	7	04/05/2011	Крестор табл. 10 мг x 28 бр.	0.000	0.000	7.000
28/04/2011	7	04/05/2011	Венофер амп 100 мг 5 мл	0.000	0.000	7.000

Начало Дата и час    Схема на приложение    Бр. Дни    Край

Общо необходимо количество    Наличност в пациента

Количество в аптечна наличност: 0.000    0.000

Количество в лекарствени единици: 0.000    0.000

Показатели    Отрисовка

# Автоматично извличане на МКБ-10 диагнози от клиничен текст

- Централна задача напр. в САЩ, наречена *автоматично кодиране*
- Болниците (Mayo Clinic) поддържат звена от мед. информатици вкл. комп. Лингвисти
- Secondary use of patient health records
- Диагнозата може да се намира навсякъде в текста (при нас – установените диагнози са в първата секция)
- Постигната голяма точност за корпуса

# БГ-Ресурси – МКБ и tabular index

- МКБ-10: 14439 описания 'код' - 'текст'

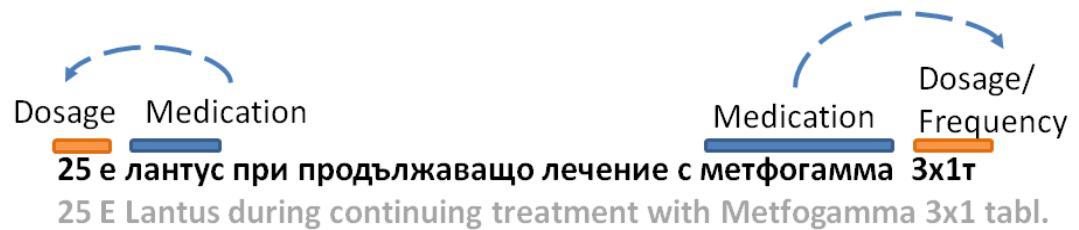
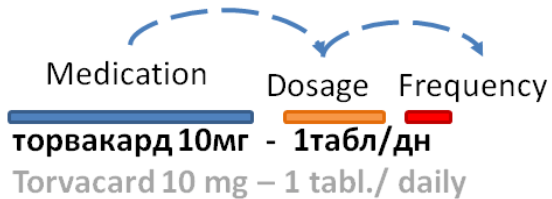
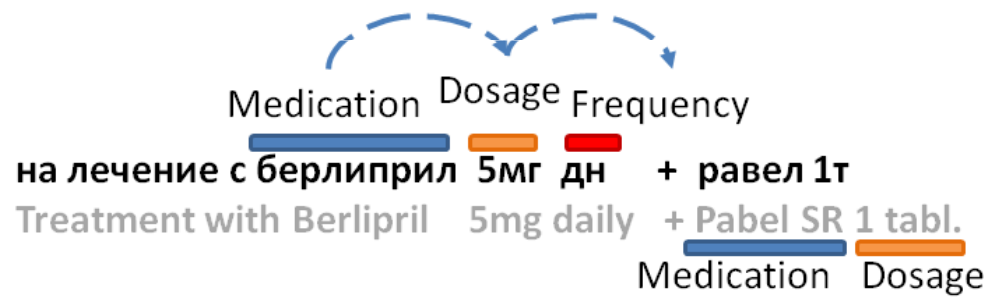
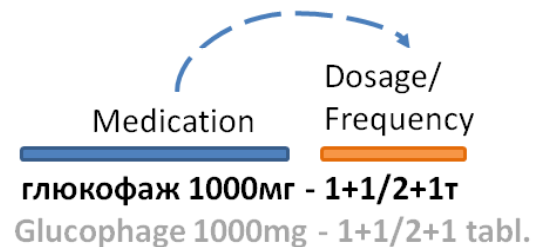
## E06.0 Остър тиреодит

- 'Указания' за прилагането ѝ

- 19 161 БГ-думи
- 291 116 БГ-словоформи
- 2 221 лат. терм. (11.59%)
- 83 713 срещания  
лат. терм. (28.76%)
- 76 939 описания на  
9044 кода от МКБ-10

**Морбили (черно) (хеморагично)  
(супресирано) B05.9**  
– германско (*виж също* Рубеола) B06.9  
– с  
– – енцефалит B05.0† G05.1\*  
– – интестинални усложнения B05.4†  
K93.8\*  
– – кератит (кератоконюнктивит) B05.8†  
H19.2\*  
– – менингит B05.1† G02.0\*  
– – отитис медия B05.3† H67.1\*  
– – пневмония B05.2† J17.1\*  
– – усложнения, НКД B05.8  
**Мориа F07.0**  
**Морон (IQ 50–69) F70.–**  
**Морска болест T75.3**

# 1500 Лекарства, АТС кодове – ОБУЧЕНИЕ с цел извличане на правила



# Резултати за 6200 епикризи

	Брой срещания в 6200 записа	Точност	Покриваемост	F Хармонично средно
Диагнози	26 826	97.30%	74.69%	84.50%
Лекарства	160 892	97.28%	99.59%	98.42%

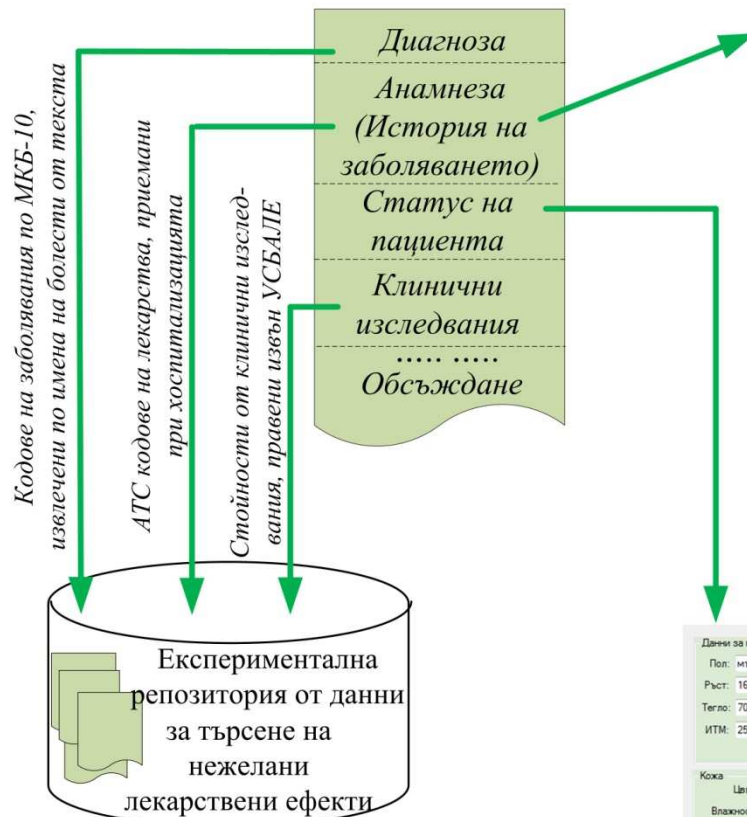


# Разпознаване на лекарства, приемани в момента на хоспитализацията (ден 0)

- Контекстуализация
- Извличане само от анамнезата при срещане на подходящи изрази
- 500 лекарства в 6200 епикризи

	Точност	Покрива-емост	F Хармонично средно
лекарства ден 0	88%	92,45%	90,17%

# 6200 анонимизи- рани епикризи



Дата(Н)	Брой Дни	Дата(К)	Наименование	Кол. от Аптека	Кол. от Клиника	Кол. от Пациент
28/04/2011	7	04/05/2011	Инсулаглар НМ 4 IU 10Ммл	0.000	0.000	1.000
28/04/2011	7	04/05/2011	Рокатрол капсул. 0.25мкг 100 бр.	0.000	0.000	7.000
28/04/2011	7	04/05/2011	Беталок табл. 50 мг	0.000	0.000	12.000
28/04/2011	7	04/05/2011	Крестор табл. 10 мг x 28 бр.	0.000	0.000	7.000
28/04/2011	7	04/05/2011	Венофер амп 100 мг 5 мл	0.000	0.000	7.000

**Record ID: 100046-08**

# Репозитория на проекта ПСИП

- Случай 26137
- Добавяне на диагнози

В БИС: E66.8 '*Other obesity*'

Извлечени от текста диагнози на придружаващи заболявания:

E89.8 '*Postprocedural adrenocortical (-medullary) hypofunction*'

E28.9 '*Other ovarian dysfunction*'

- Лекарства, стойности от клинични изследвания – вж.файл

# КОЛЕКТИВ

- Димитър Чаръкчиев
- Галя Ангелова
- Светла Бойчева
- Ивелина Николова
- Христо Димитров
- Живко Ангелов

**БЛАГОДАРИМ ВИ ЗА ВНИМАНИЕТО!**

**Следва демонстрация в реално време**