

Tecnologies del Llenguatge aplicades a la Sanitat

Com analitzar de forma automàtica documentació mèdica estructurada i no estructurada per generar coneixement mitjançant l'ús de processament del Llenguatge Natural.

Montse Cuadros
mcuadros@vicomtech.org

Test

Què són les tecnologies del llenguatge?

- On s'apliquen?
- Què és important saber?
- Beneficis

Tecnologies del llenguatge aplicades a la salut

Explotació text mèdic

Què són les tecnologies del llenguatge?

Què són les tecnologies del llenguatge?

- o El camp de coneixement s'anomena PLN (Processament del Llenguatge Natural) i es una disciplina que combina Intel·ligència Artificial i Lingüística Computacional.
- o Les tecnologies del llenguatge permeten que les màquines processin, entenguin i interpretin el llenguatge natural, tan si es escrit com si es parlat.
- o Tres nivells d'anàlisi:
 - Sintàctic: Quina estructura té el text.
 - Semàntic: Que vol dir el text.
 - Pragmàtic: Que vol realment dir el text.
- o Es tenen en compte fenòmens com la fonètica i la morfologia.

Què són les tecnologies del llenguatge? ntes

Tecnologies del llenguatge natural parlat

- Reconeixement de la veu
- Síntesis de veu

Tecnologies del llenguatge natural escrit

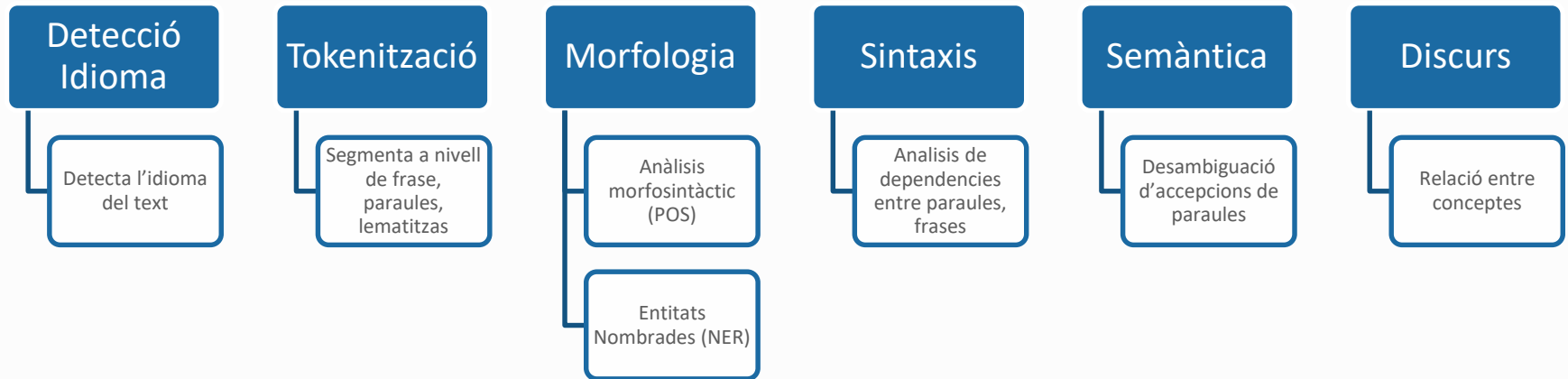
- Correcció ortogràfica/gramatical
- Extracció de terminologia
- Detecció de opinions/sentiments
- Búsqueda i extracció de continguts
- Traducció automàtica

Tecnologies interacció home-màquina

- Comprensió de llenguatge parlat
- Gestió del diàleg
- Generació de llenguatge natural

Què són les tecnologies del llenguatge?

A nivel de text, els passos que normalment es segueixen per analitzar-lo són els següents:



Què són les tecnologies del llenguatge?

Correu electrònic

Assumpte: ponència a la UHC

Data: 02-02-2018

Destinatari: Montse Cuadros

Hola Montse,

Et recordo que dilluns has de ser a Barcelona, C/València 333, a la seu de la Unió de Hospitals de Catalunya per fer una ponència entre les 10 i les 13.

Fins aviat,
Maria



Acte: ponència a la Unió de Hospitals de Catalunya

Data: 05-02-2018

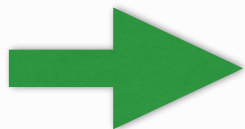
Hora inici: 10:00

Hora final: 13:00

Lloc: C/València 333,
Barcelona

Què són les tecnologies del llenguatge?

Yesterday I went to a nice beach in Barcelona where I saw my friend Joan.



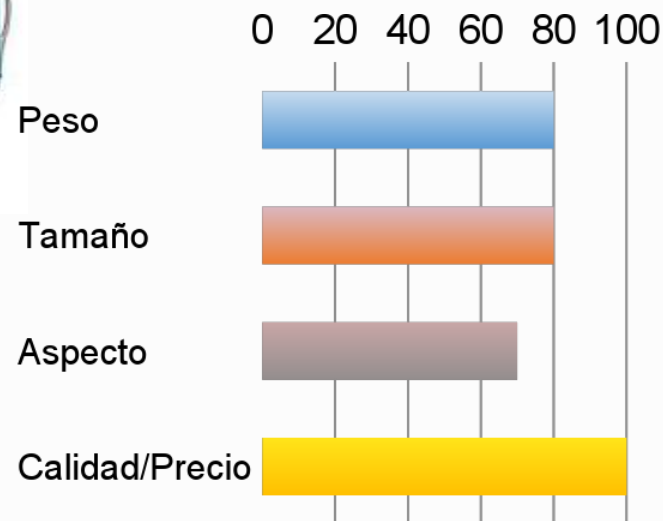
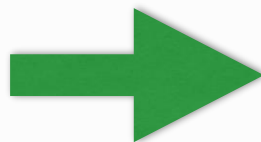
Ahir vaig estar en una platja molt bonica de Barcelona y vaig veure el meu amic amic Joan.

Què són les tecnologies del llenguatge?

Maca i fàcil de portar

Al **pesar poc** i **ser petita**, no he de portar la réflex. A més, **qualitat-preu es fantàstica**.

La càmera té un tacte plasticós, es **molt lleugera** pero sembla que es podria trencarà.



On s'apliquen?

Sector públic/administratiu

Analizar documentació administrativa

Ordenar i classificar documentació

Traduir automàticament

Generar informes

Sector de la publicitat i marketing

Extreure el sentiment i l'opinió dels clients respecte a les empreses

Buscar intenció de compra dels clients

Sector turístic

Extreure el sentiment i la opinió dels clients respecte a hotels i/o allotjaments en portals web

Aplicaciones de las tecnologías de la lengua

On s'apliquen?

Sector de la traducció

Augmentar la productivitat dels traductors a través de la post-edició de traduccions automàtiques

Sector audiovisual

Augmentar la productivitat del procés de subtitulació/transcripció a través de la post-edició de subtítulos/transcripciones automáticas

Sector de la educació

Donar suport al professorat en la generació de material per l'educació

Traducció automàtica dels MOOC (Massive Online Open Courses)

Aplicaciones de las tecnologías de la lengua

On s'apliquen?

Sector de la fabricació

Suport en la escriptura de manuals tècnics

Traducció automàtica de manuals tècnics

Sector de les tecnologies de la informació

Automatització de l'atenció al client

Categorització automàtica de factures

Sector salut

Anàlisis d'historials mèdics per donar suport a la decisió

Traducció automàtica per superar les barreres lingüístiques d'atenció a pacients

Suport a la codificació CIE-10

Què és important saber?

- És una tasca difícil.
- El llenguatge es ambigu i evoluciona amb el temps.
- Diferents anàlisis possibles, significats, ..
- Coneixement especialitzat (del idioma, fonts de coneixement, tècniques aprenentatge automàtic).
- Depèn:
 - del idioma del text / audio (anglès, català, castellà, basc, xinès, àrab)
 - del domini (jurídic, mèdic, màquina-herramienta)
 - del origen (texts de sms, informes mèdics, texts periodistics, audios ...)
 - volum de dades

Beneficis

- Automatitzar anàlisis de continguts
- Millorar processos mitjançant l'automatització de molts anàlisis
- Obtenir resultats de forma més ràpida que de la forma tradicional.
- Treure valor als continguts dels organismes, que creixen exponencialment al llarg del temps

Tecnologies del llenguatge aplicades a la salut

Peculiaritats del domini (volum)

Gran volum de dades

179,5 milions de documents clínics electrònics

~ 80% dels documents consisteix en text llibre

% de text estructurat

% de text no estructurat

Peculiaritats del domini (estructura)

Text no estructurat:

- Informes mèdics
- Articles científics
- RIS/PACS
- HCE
- Dades de monitorització
- Dades òmiques
- open data
- Entrades de pacients
- Guies clíniques
- Foros de salut online
- Publicacions científiques
- etc.

Text estructurat

- Bases de dades de medicines, fàrmacs.
- Camps en informes, formularis mèdics uníbocs

Peculiaritats del domini (llenguatge)

Estructures sintàctiques variables:

- Abdomen: Blando y depresible
- T.A:160/106 mmhg. F.C:74x'. Ta:36'1o.
- adenopatías de tamaño patológico En parénquimas pulmonares se aprecian áreas de condensación

Gramàtica atípica:

- No [se aprecia] Hernia de Hiato
- Eupneica en reposo

Llenguatge mèdic:

- adherencias de la IQ. Foco infeccioso en LSD
- no palpo masas ni megalias

Peculiaritats del domini

Abreviatures, Errors tipogràfics

Ambiguitat:

- Depen de la zona geogràfica o ús
- Polisèmia

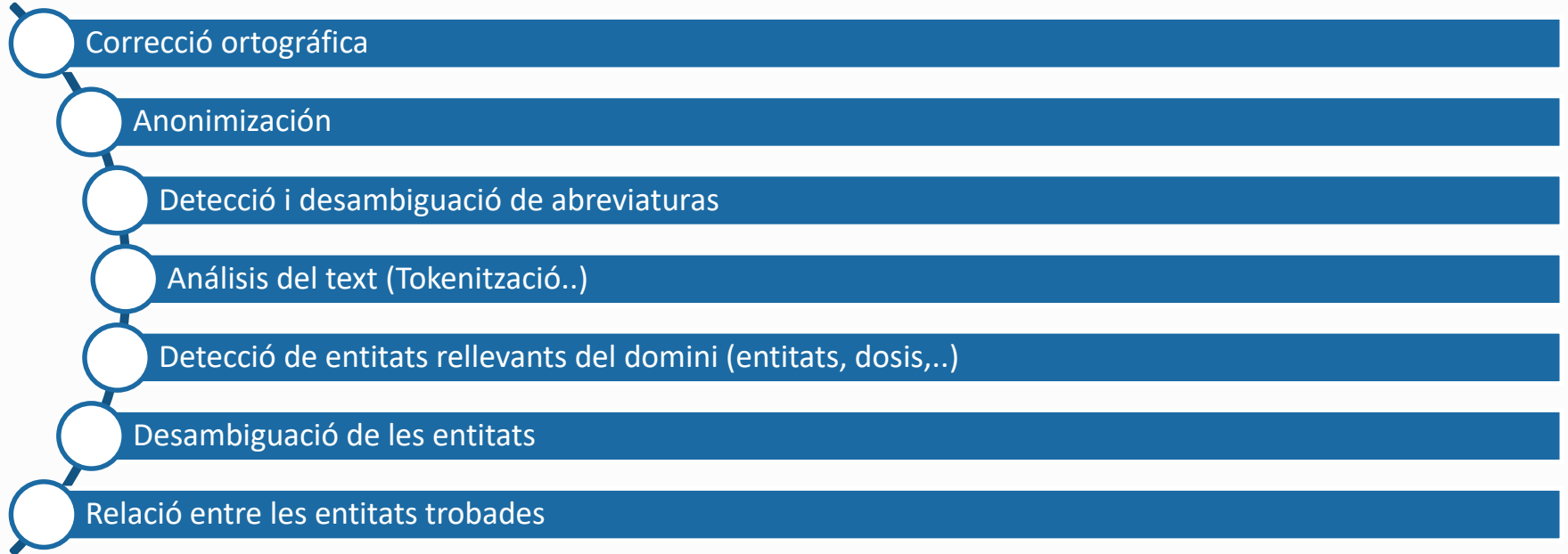
Variabilitat lèxica:

- Un mateix concepte es pot escriure de diferents maneres (UMLS)

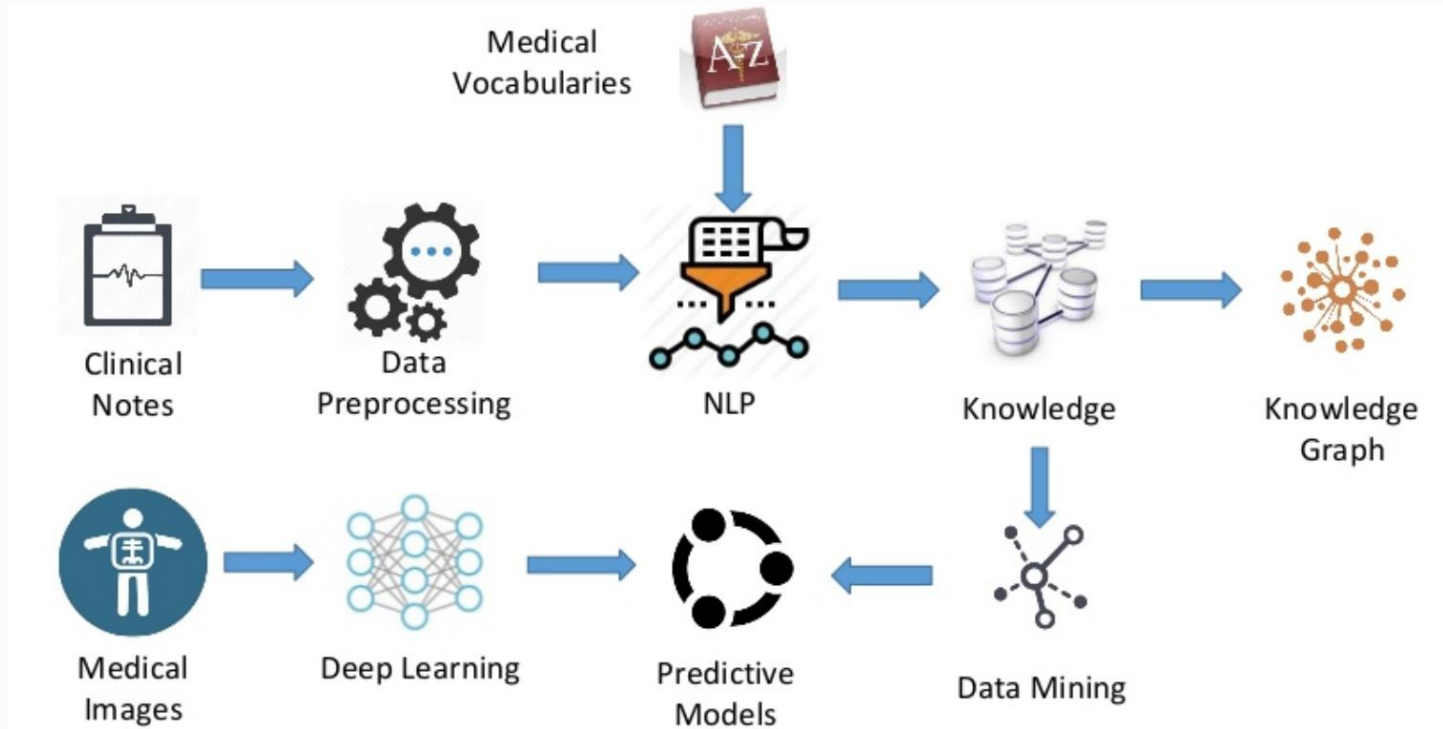
[C0015417]
protesis de globo ocular
protesis ocular
ojo artificial
ojo protesico
implante ocular esferico

[C0019202]
enfermedad de Wilson
degeneracion neurohepatica
pseudoesclerosis cerebral
sndrome de Westphal-Strumpell
(+11)

Pasos a seguir en el processament del text mèdic



Anàlisi del text mèdic



Demo

Test

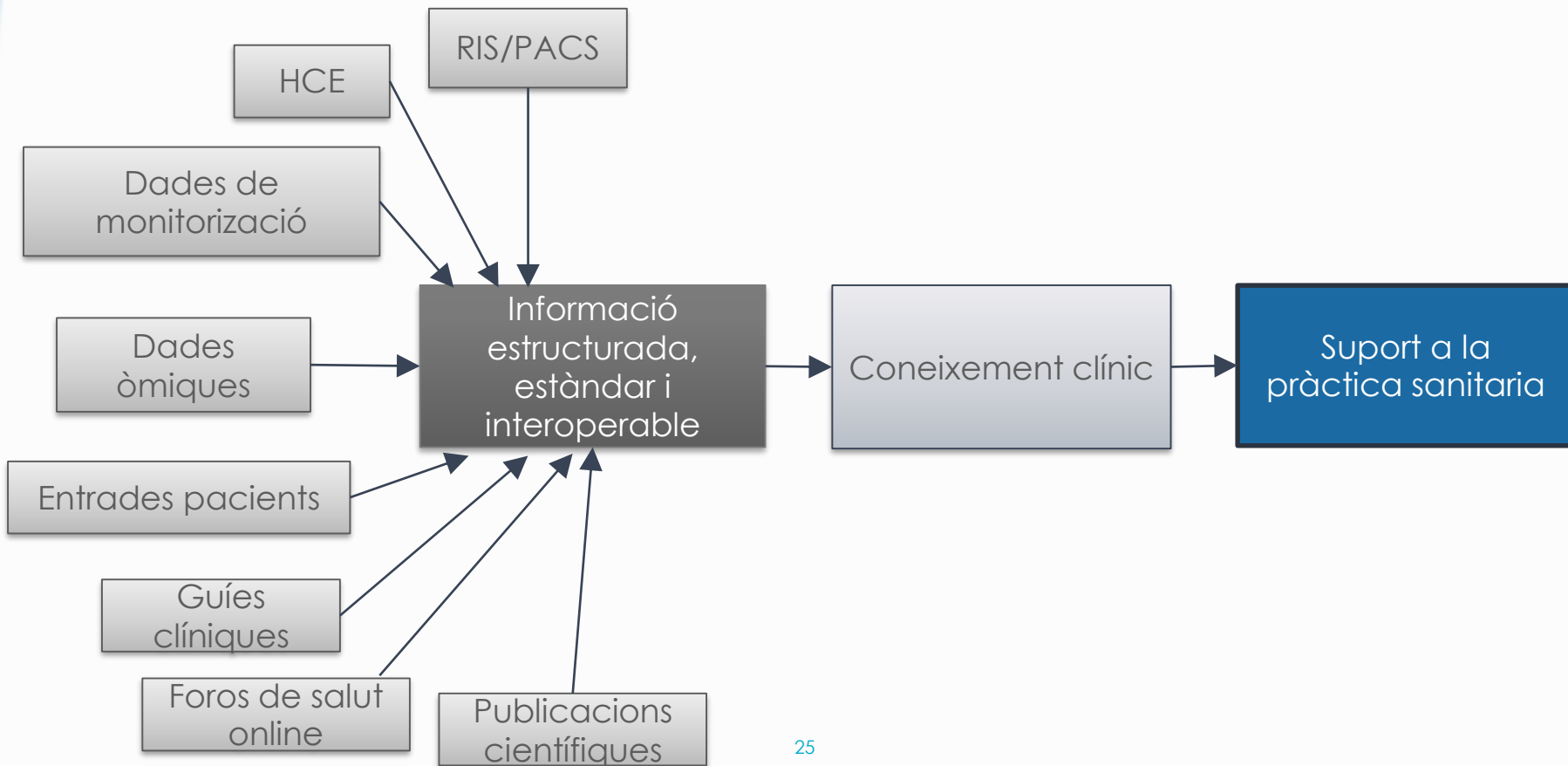
[UMLSmapper](#)

Explotació text mèdic

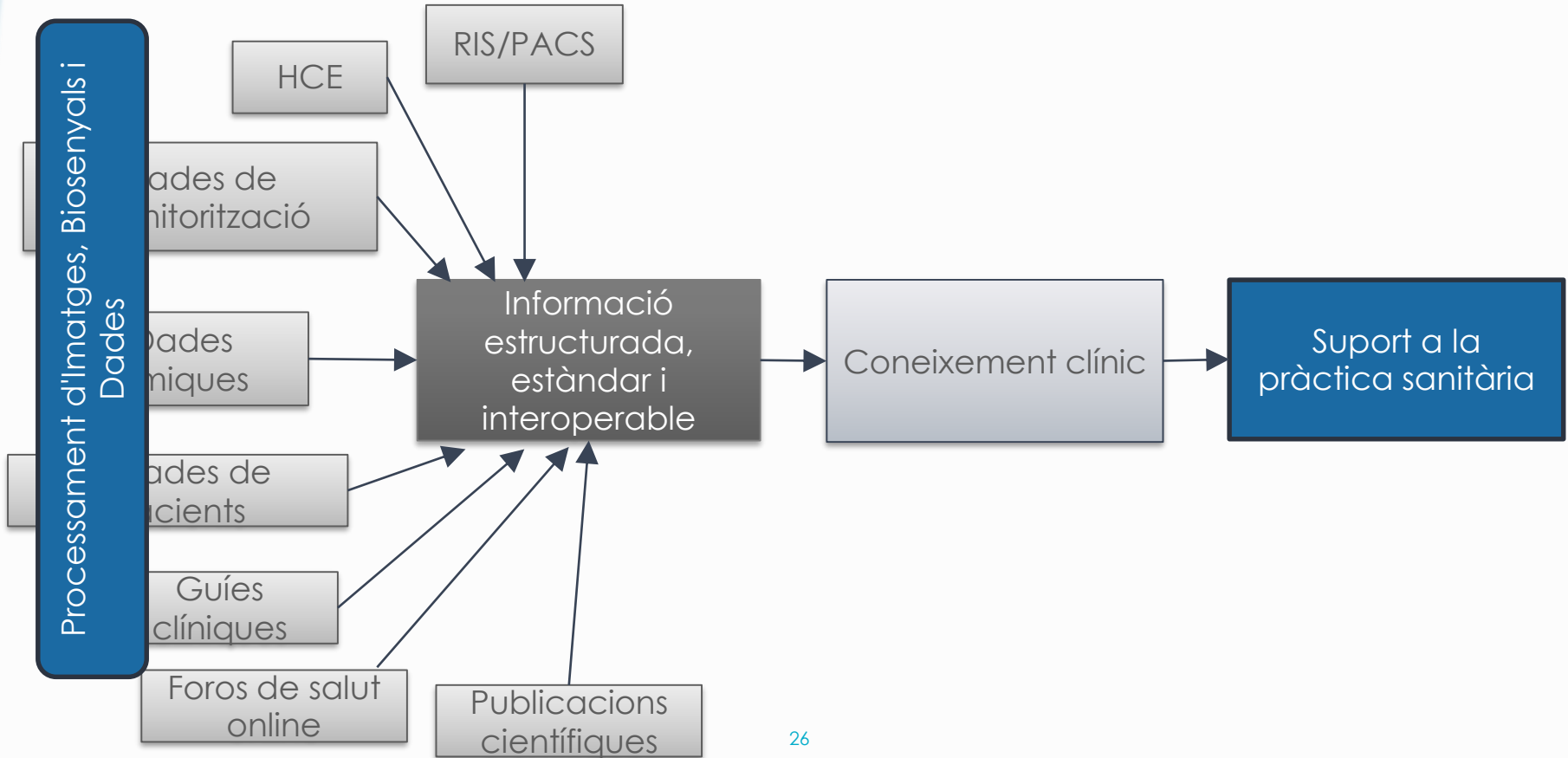
Objectiu

- Proveir al sector sanitari de tecnologia per explotar de forma més eficient les dades estructurades i no estructurades presents en els sistemes d'Informació sanitària.

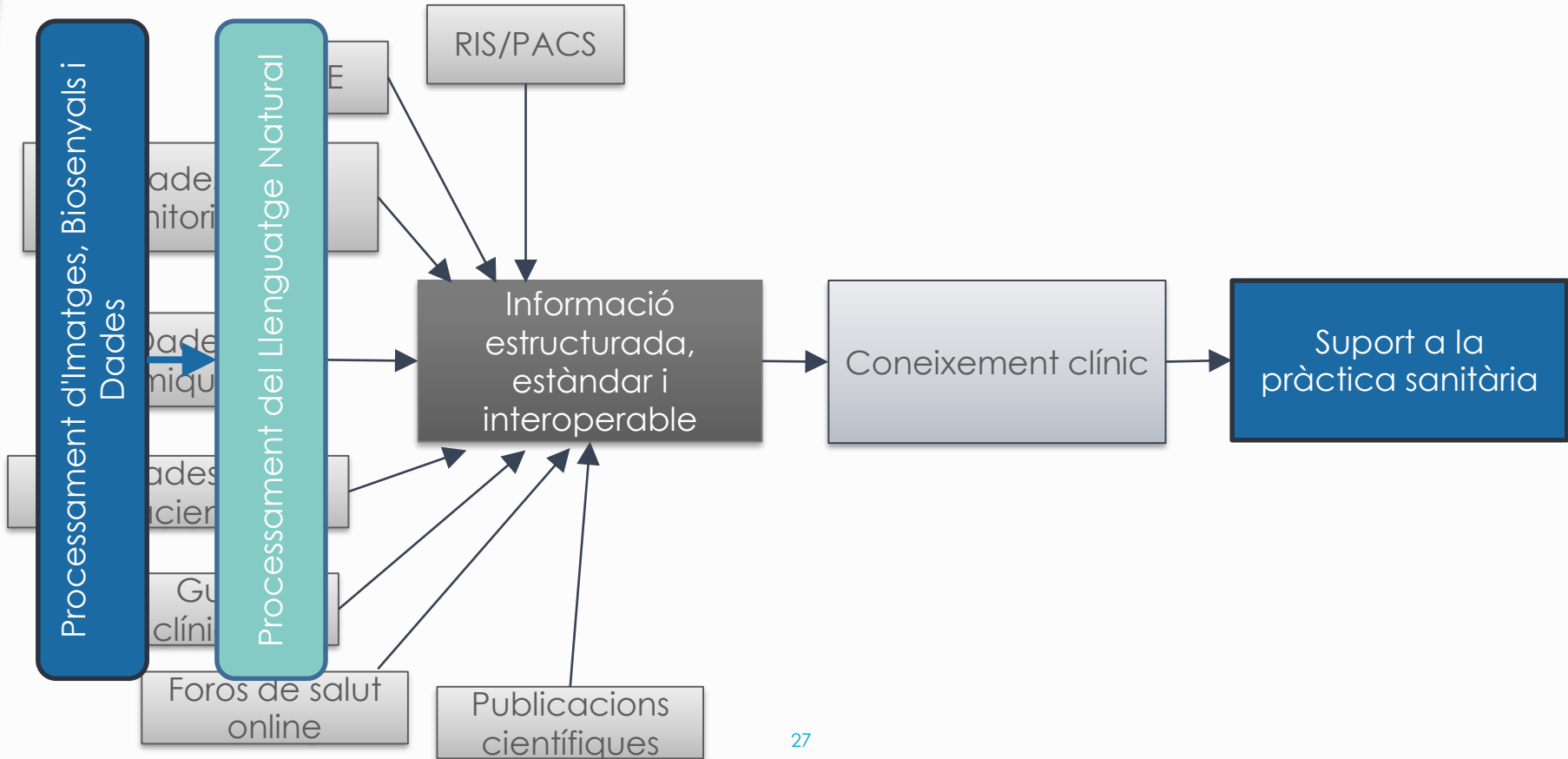
Objectiu



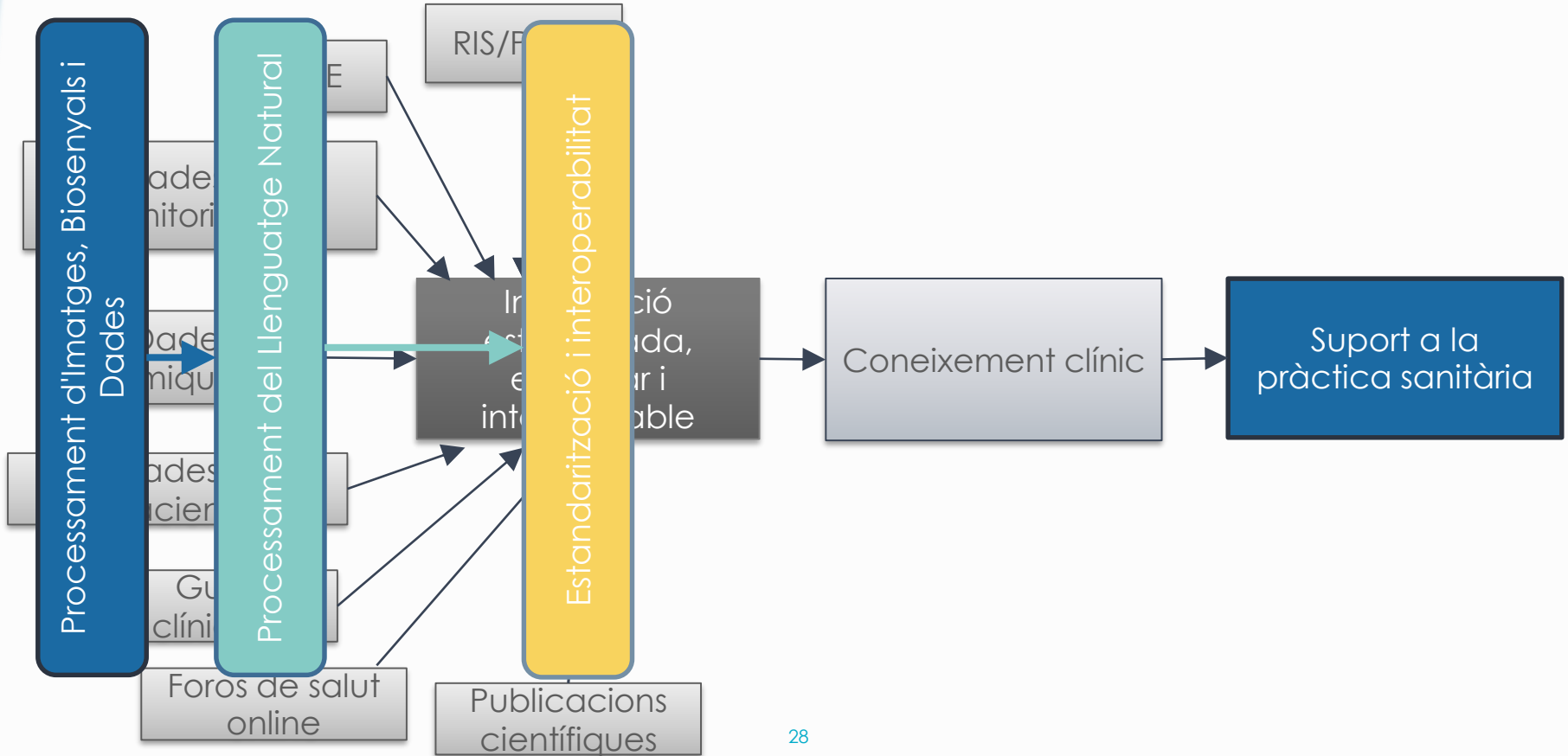
Línies d'investigació



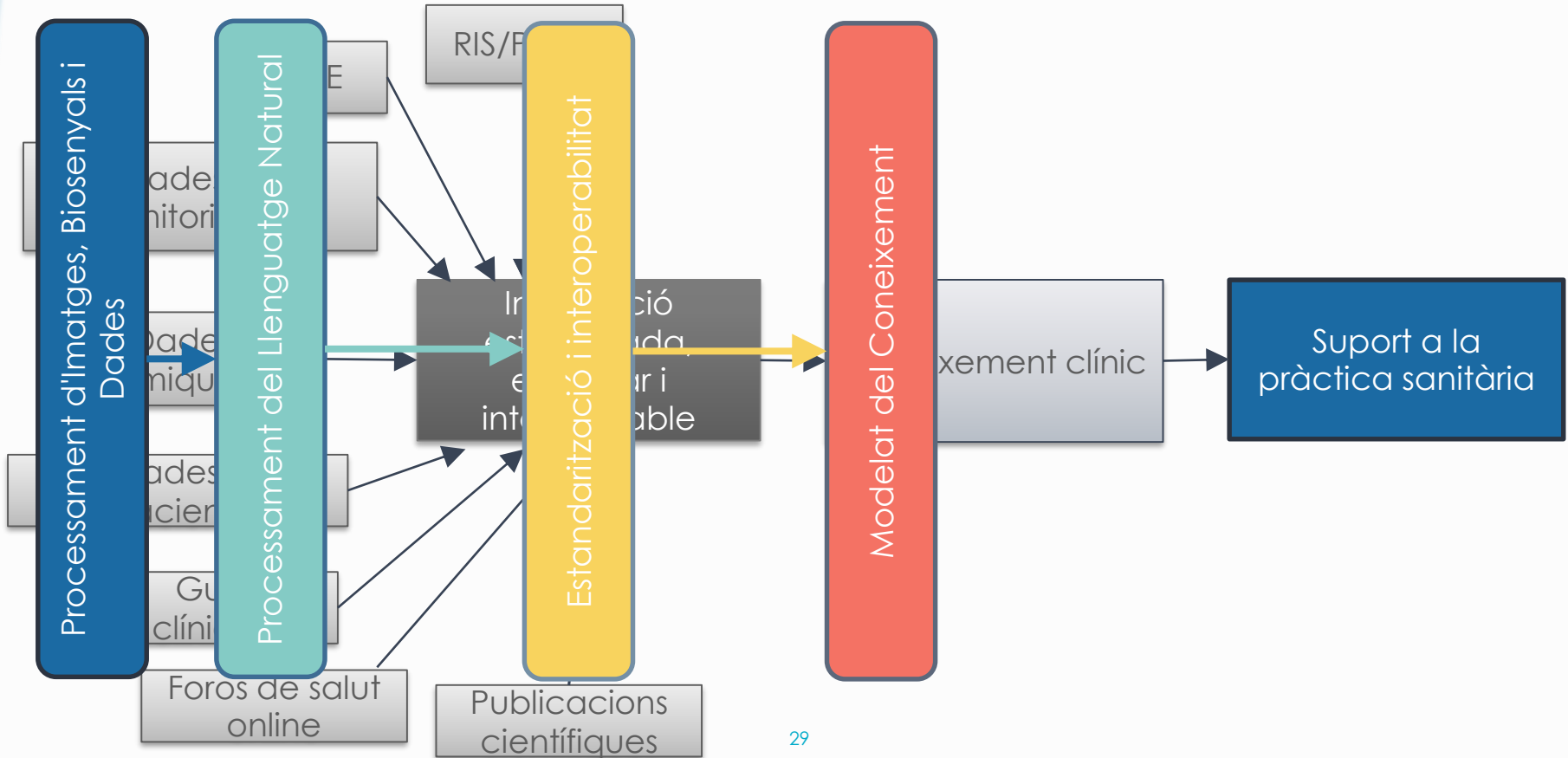
Línies d'investigació



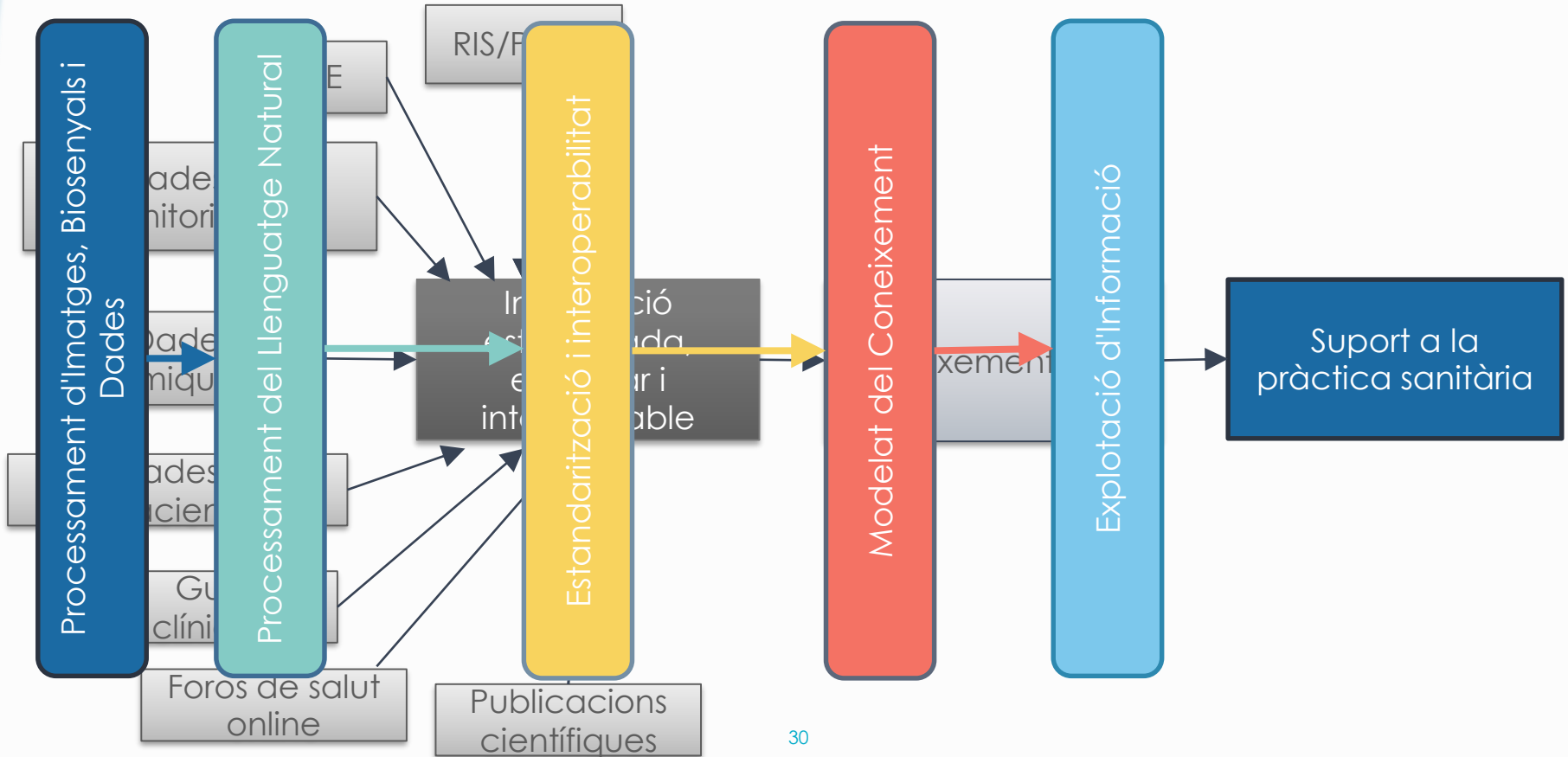
Línies d'investigació



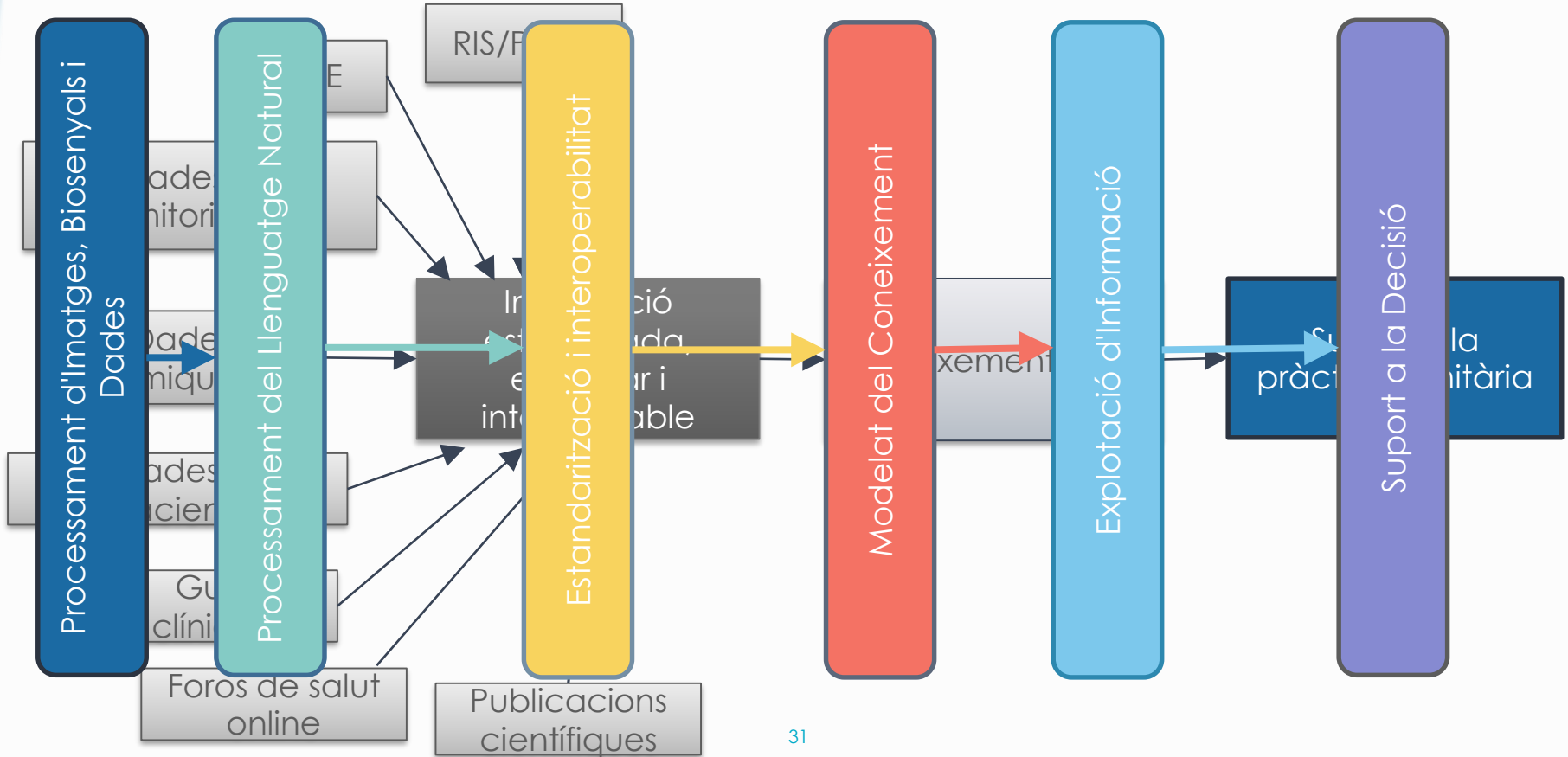
Línies d'investigació



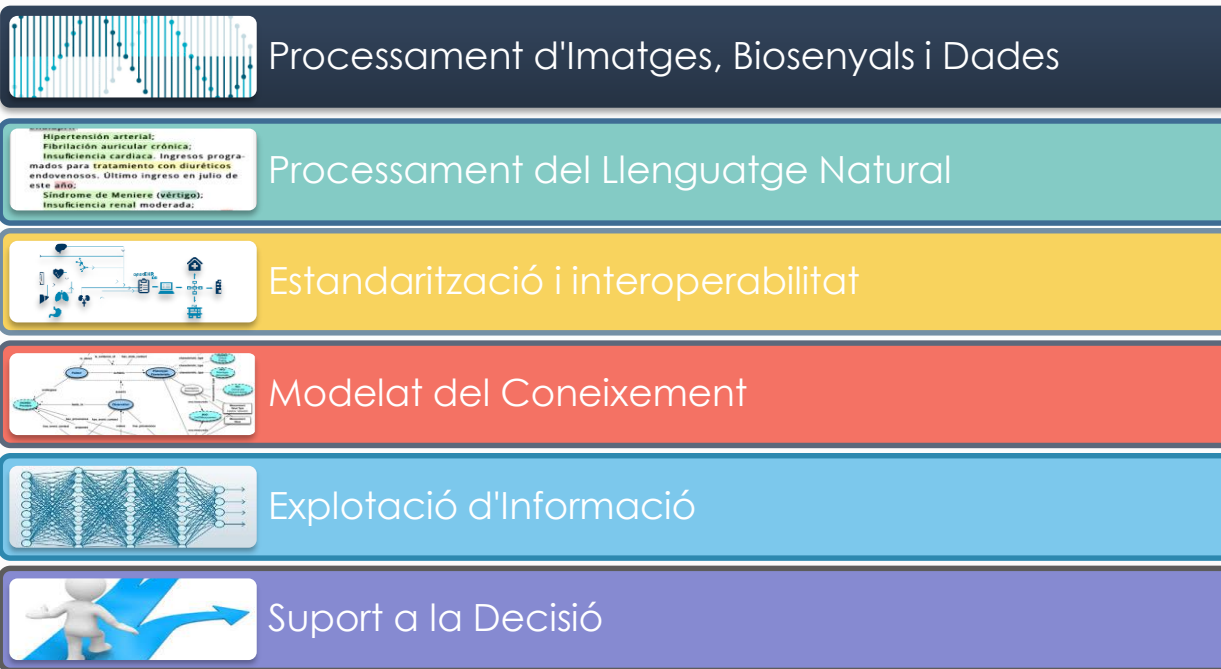
Línies d'investigació

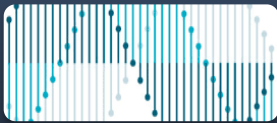


Línies d'investigació



Línies d'investigació





- Anonimització i conformitat legal
- Processament de imatge
 - Normalització i filtratge
 - Extracció de característiques i modelat
 - Segmentació i anotació
- Captura i pre-processat de biosenyals
 - Wearables i Plataformes Online de Salut (p.ej. GoogleFit, AppleHealth)
 - Dispositiu mèdic al costat del llit
 - Integració a dispositiu a través d'estandards (p.ej. Continua Alliance)
- Filtrat de dades
 - Qualitat de les dades
 - Filtrat, homogenització

Processament del Llenguatge Natural

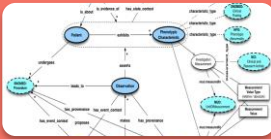
Hipertensió arterial;
Fibril·lació auricular crònica;
Insuficiència cardíaca. Ingressos programats para tratamiento con diuréticos endovenosos. Último ingreso en julio de este año;
Síndrome de Meniere (vértigo);
Insuficiencia renal moderada;

- Dictat i transcripció automàtica d'informes
- Detecció i desambiguació d'acrònims mèdics
- Detecció d'entitats nombrades (síntomes, malalties, òrgans i parts del cos, medicaments, etc.)
- Recuperació d'informació i búsqueda de respostes
- Generació automàtica de resums
- Síntesis d'informes a partir de dades
- Sistemes de diàleg i bots



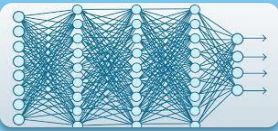
Estandarització i interoperabilitat

- Capa d'interoperabilitat necessària
- Ús i aplicació d'estandars existents i eines associades
 - Estàndars semàntics
 - RDF
 - NCI Thesaurus
 - SNOMED CT
 - CIE 10
 - UMLS
 - Estàndars sintàctics
 - FHIR (HL7)
 - Estàndars HCE
 - OpenEHR
 - ISO13606
 - HL7
- Aportació puntual als processos d'estandarització



Modelat del Coneixement

- Formalització interpretable per el ordenador del coneixement associat a un domini o patologia i que prove de diverses fonts (Guies clíniques, publicacions, etc.)
 - Model Digital del Pacient
 - Medicina Basada en la Evidencia (p.ej. from Clinical Practice Guidelines to Computer Interpretable Guidelines)
 - Modelat de la Experiencia Clínica



- Aplicació de diverses tècniques d'aprenentatge automàtic
 - Models predictius
 - Clasificació/Estratificació
 - Tècniques de raonament intel·ligent (case-based, experience-based, etc.)
 - Machine learning
 - Deep learning
- Big Data
- Outcome analysis



- Interacció amb l'usuari final (pacient, clínic, gestor) per facilitar la presa de decisions
- Sistemes de Suport a la Decisió Clínica (p.ej. basats en guies clíniques, basats en casos, basat en experiència)
- Sistemes de Soport per el Pacient
 - Guiat del Pacient en situacions de mobilitat
 - Empoderamiento de Pacient
 - Asistente Virtual de Salut
- Quadres de Mando, Analítica Visual, Business Intelligence
- Visualització avançada d'imatge

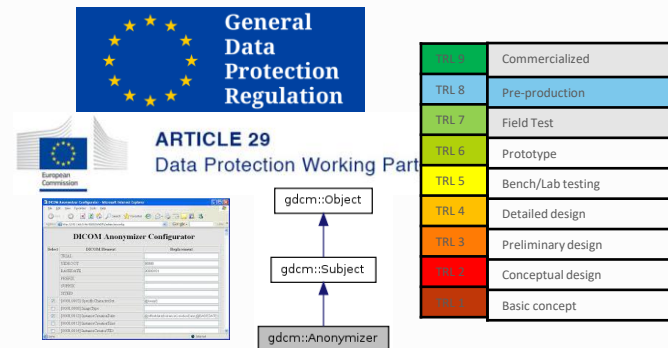
Aplicacions

- PER PROFESIONALES
 - Anonimització
 - Codificació automàtica i en continu
 - Smart EHR
 - Suport a la redacció ràpida i precisa
 - Suport a la Decisió clínica
 - Modelat i estratificació de pacients en medicina personalitzada
 - Búsqueda d'informació i evidència en fonts contrastades
 - Farmacovigilància
 - Radiología cognitiva



Anonimizació

- Existeixen un conjunt de normatives (p.ej. **GDPR, LOPD**), recomanacions (p.ej. **Art29WP opinion on anonymization techniques**) i eines (p.ej. **CTP DICOM Anonymizer, GDCM**) per anonimitza dades de caràcter personal en diferents àmbits d'aplicació.



A Vicomtech-IK4 **fem servir i apliquem** els **mecanismes** d'anonimizació **existents** i **desenvolupem tècniques** per fer que el **procès** d'anonimizació sigui **més productiu**.

RADMOVE
Anonimizació de radiografias

ASGARD
Anonimizació de missatges en xarxes
socials
H2020

IRTEN
Anonimizació de trucades de atenció al
client

L. García-Sardiñas, M. Serras and A. del Pozo. **ES-Port: a Spontaneous Spoken Human-Human Technical Support Corpus for Dialogue Research in Spanish**. Language Resources and Evaluation Conference, 2018 (submitted)

Suport a la redacció ràpida i precisa

- Els HIS/EHR no inclouen eines de suport a la escriptura adaptades al context de cada servei sanitari



TRL 9	Commercialized
TRL 8	Pre-production
TRL 7	Field Test
TRL 6	Prototype
TRL 5	Bench/Lab testing
TRL 4	Detailed design
TRL 3	Preliminary design
TRL 2	Conceptual design
TRL 1	Basic concept

Vicomtech-IK4 aporta tota la **cadena de valor** per què la **historia clínica** escrita s'inclogui **de la forma més àgil i còmode** per el professional i amb el **mínim nombre d'errors**:

- **agilizar** la **escriptura** (escriure més ràpid)
- **detectar errors** ortogràfics i gramaticals i **oferir correccions en temps real**
- integrar **diccionaris mèdics** generals/específics
- integrar **diccionaris de sigles mèdiques**

escriMED

Apoyo a la escritura médica y
codificación automática de
informes médicos

Codificació automàtica CIE-10

- Hi ha varies **empreses** al **mercat** amb **solucions** de **codificació** de **diferent naturalesa**



TRL 9	Commercialized
TRL 8	Pre-production
TRL 7	Field Test
TRL 6	Prototype
TRL 5	Bench/Lab testing
TRL 4	Detailed design
TRL 3	Preliminary design
TRL 2	Conceptual design
TRL 1	Basic concept

Vicomtech-IK4 aporta les **últimes millores** en I+D del sector **per adaptar-les** a les **necessitats** concretes de **cada client**

Estructuració de text mèdic i en contínu

- En general, **avui en dia** només s'explota la informació estructurada de la HCE i **no s'explota la informació guardada en text lliure, malgrat que el 80% de la HCE ho sigui. Algunes empreses** estan **començant a oferir eines** per extreure i explotar informació estructurada de la HCE.



TRL 9	Commercialized
TRL 8	Pre-production
TRL 7	Field Test
TRL 6	Prototype
TRL 5	Bench/Lab testing
TRL 4	Detailed design
TRL 3	Preliminary design
TRL 2	Conceptual design
TRL 1	Basic concept

Vicomtech-IK4 disposa de **tecnologia madura** que **estructura de forma automàtica text lliure sanitari basant-se en terminologies internacionals estandaritzades** (SNOMED CT, UMLS) i que permeten la seva posterior explotació mitjançant quadros de mando, tècniques d'anàlítica visual o sistemes de Suport a la Decisió.

SemanHIS

Semantització de Història Clínica
Electrònica

N. Pérez, “**Mapping of Electronic Health Records in Spanish to the Unified Medical Language System Metathesarus**”, Tesis de Máster Análisis y Procesamiento del Lenguaje, UPV/EHU, 2017

N. Pérez, M. Cuadros and G. Rigau “**Annotation of Spanish electronic health records with UMLS**”, TUNER Workshop, SEPLN, 2017

N. Pérez, M. Cuadros and G. Rigau “**Biomedical term normalization of EHRs with UMLS**”, LREC 2018

Smart EHR

- Les **HIS/EHR** estan **evolucionant** ràpidament i estan **incorporant** un gran nombre de **tecnologies** heterogènies.



TRL 9	Commercialized
TRL 8	Pre-production
TRL 7	Field Test
TRL 6	Prototype
TRL 5	Bench/Lab testing
TRL 4	Detailed design
TRL 3	Preliminary design
TRL 2	Conceptual design
TRL 1	Basic concept

Vicomtech-IK4 **aporta** les **últimes tecnologies** d'**estructuració, explotació i visualització** de **dades** als HIS/HER per:

- **millorar l'eficiència sanitària** i l'**atenció al pacient**
- **automatitzar** els **procesos** sanitaris
- **explotar** la **informació** de forma **intelligent** per **àmbit sanitari, predir automàticament events adversos** (p.ej. reingresos), etc.

INCAR
Sistema de monitorización y guiado
Inteligente para pacientes con
insuficiencia CARDíaca

SmartHealth
Plataforma para el Suport a la
Decisió Clínica y la Monitorización
de la Salud

SemanHIS
Semantización de Historia Clínica
Electrónica

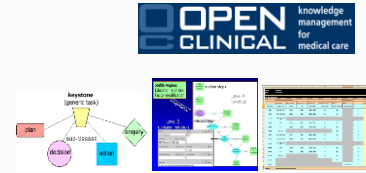
Muro N., et al. "**Hygehos Ontology for Electronic Health Records**". InMed 2016: Innovation in Medicine and Healthcare 2016 pp 311-321.

Murua A., et al. "**Upgrading legacy EHR systems to Smart EHR systems**". InMed 2017: Innovation in Medicine and Healthcare 2017 pp 227-233.

Artetxe A. "**Computational Intelligence Contributions to Readmission Risk Prediction in Healthcare Systems**". Doctoral Thesis, UPV 2017.

Suport a la Decisió clínica

- Existeix un conjunt de **mètodes** (p.ej. Asbru, PROforma, GLIF3) i **eines** (p.ej. Drools) per el **desenvolupament de guies clíniques interpretables per l'ordenador**.



TRL 8	Commercialized
TRL 8	Pre-production
TRL 7	Field Test
TRL 6	Prototype
TRL 5	Bench/Lab testing
TRL 4	Detailed design
TRL 3	Preliminary design
TRL 2	Conceptual design
TRL 1	Basic concept

Vicomtech-IK4 esta a la **vanguardia** en **investigació** per **incloure** a més de guies clíniques l'**experiencia clínica en sistemes de Suport a la Decisió**.

LIFE

Desafío integral al cáncer de mama

Desiree

Decision Support and Information Management System for Breast Cancer H2020

N. Larburu, N. Muro, I. Macía, E. Sánchez, H. Wang, J. Winder, J. Boaud and B. Séroussi. **Augmenting Guideline-based CDSS with Experts' Knowledge**. 10th International Joint Conference on Biomedical Engineering Systems and Technologies, (HEALTHINF, BIOSTEC-17) 21-23 February 2017, Porto, Portugal.

N. Muro, N. Larburu, J. Bouaud, J. Belloso, G. Cajaraville, A. Urruticoechea, and B. Séroussi. **Augmenting Guideline Knowledge with Non-Compliant Clinical Decisions: Experience-Based Decision Support**. 5th International Conference on Innovation in Medicine and Healthcare (InMed-17) 21-23 June 2017, Algarve, Portugal.

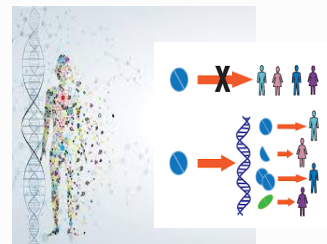
B. Seroussi, G. Guézennec, J. B. Lamy, N. Muro, N. Larburu, B. Sekar and J. Bouaud. **Reconciliation of multiple guidelines for decision support: a case study on the multidisciplinary management of breast cancer within the DESIREE Project**. American Medical Informatics Association 2017 Annual Symposium (2017 AMIA) 4-8 November 2017, Washington D.C., USA.

N. Muro. **Development of an Advanced Clinical Decision Support System: Enriching the Guideline-Based Knowledge with Experience**. 16th Conference on Artificial Intelligence in Medicine (AIME 2017) 21-24 June 2017, Vienna, Austria.

N. Muro, N. Larburu, J. Bouaud and B. Séroussi. **Weighting Experience-Based Decision Support on the Basis of Clinical Outcomes' Assessment**. EFMI STC 2017: "The practice of patient centered care: Empowering and engaging patients in the digital era". 22-23 October 2017, Tel Aviv, Israel.

Modelat i estratificació de pacients en medicina personalitzada

- El paradigma de la **Medicina Personalitzada** exigeix desenvolupar **tècniques de perfilat** de los **pacients** per poder recomandar la terapia més adequada a les característiques personals.



	Commercialized
TRL 8	Pre-production
TRL 7	Field Test
TRL 6	Prototype
TRL 5	Bench/Lab testing
TRL 4	Detailed design
TRL 3	Preliminary design
TRL 2	Conceptual design
TRL 1	Basic concept

Vicomtech-IK4 es **expert** en tècniques de **clustering** i **machine learning** per l'**estratificació de pacients**.

IMPETUS
Tecnologies para el Impulso de
la Medicina Personalizada para
su Uso clínico

Artetxe A., et al. "**Activity Classification Using Mobile Phone based Motion Sensing and Distributed Computing**". InMed14: Innovation in Medicine and Healthcare 2014.

Sánchez R., et al. "**A platform for user empowerment through Self Ecological Momentary Assessment / Intervention**". 5th EAI International Conference on Wireless Mobile Communication and Healthcare, 2015.

- PER INVESTIGACIÓ
 - Real World Evidence/ Real World Data
 - Gestió de cohorts
 - Búsqueda de perfils per assajos clínics



Real World Evidence / Real World Data

- Els **estudis clínics** estan sotmesos a varies **limitacions** per tant els **resultats** de las **terapies** en la **pràctica real** en ocasions **són diferents** als **esperats**.



TRL 9	Commercialized
TRL 8	Pre-production
TRL 7	Field Test
TRL 6	Prototype
TRL 5	Bench/Lab testing
TRL 4	Detailed design
TRL 3	Preliminary design
TRL 2	Conceptual design
TRL 1	Basic concept

Vicomtech-IK4 té **tecnologia** capaç d'**analitzar de forma sistemàtica i automatizada** els **resultats reals** de les **terapies** en la **clínica diària** (entrades de pacients) per **extreure nou coneixement**.

IMPETUS

Tecnologies para el Impulso de la Medicina Personalizada para su Uso clínico

INCAR

Sistema de monitorización y guiado Inteligente para pacientes con insuficiencia CARDíaca

DESIREE

Decision Support and Information Management System for Breast Cancer
H2020

Sanchez E., et al. "**Decisional DNA for modeling and reuse of experiential clinical assessments in breast cancer diagnosis and treatment**". Neurocomputing, Vol. 146, pp. 308-318, December, 2014.

Muro N., et al. "**Experience-Based Electronic Health Records**". Cybernetics and Systems, Volume 47, pp. 126-139, 2016.

3 exemples concrets de projectes H2020 en curs d' Explotació d'Informació sanitaria estructurada



DESIREE Decision Support and Information Management System for Breast Cancer

– Suport a la Decisió en Comitès de Mama



MIDAS Meaningful Integration of Data, Analytics and Services

– Big Data en Serveis de Salut Pública



CAPTAIN Coach Assistant via Projected and Tangible Interface

– Assistents de Salut per persones grans a la llar

Clients i col·laboradors - estatals

Ibermática

Bilbomática 

INFORMÁTICA
El Corte Inglés

 **SISTEMAS GENÓMICOS**
ΔSCIRES

 **HYGEHOS**

 **igarle**
HEALTH

 **PlenarTech**

 **clínica Asunción klinika**


Osakidetza

bioef

biocruces
health research institute


Onkologikoa


EUSKO JAURLARITZA GOBIERNO VASCO
OSASUN SAILA
DEPARTAMENTO DE SALUD

Clients colaboradores - internacionais



Xarxes i aliances



bi+odonostia

osasun ikerketa institutua
instituto de investigación sanitaria

APTES



Equip

- Directors
 - Dr. Iván Macía, eSalud i Aplicacions Biomèdiques
 - Dra. Arantza del Pozo, Tecnologies de la Parla i el Llenguatge Natural
- Investigadors senior
 - Dr. Eduard Carrasco, HCE & Empoderamiento de Usuario
 - Dr. Luis Kabongo, Processament de Imatge Mèdica
 - Dra. Nekane Larburu, Sistemas de Suport a la Decisió
 - Dra. Montse Cuadros, Processament del Llenguatge Natural
- Investigadors
 - Dr. Arkaitz Artetxe, Models Predictius & Data Mining
 - Dr. Aitor G. Pablos, Processament del Llenguatge Natural
 - Naiara Pérez, Processament del Llenguatge Natural
 - Roberto Álvarez, Estandarizació & Interoperabilitat
- Ajudants d'investigació
 - Naiara Muro, Sistemas de Suport a la Decisió Clínica, HCE

11 investigadors

73% doctors

amb experiència a

- FP7 & H2020
- CDTI & CIEN



www.vicomtech.org
info@vicomtech.org

