Dataportal Platform

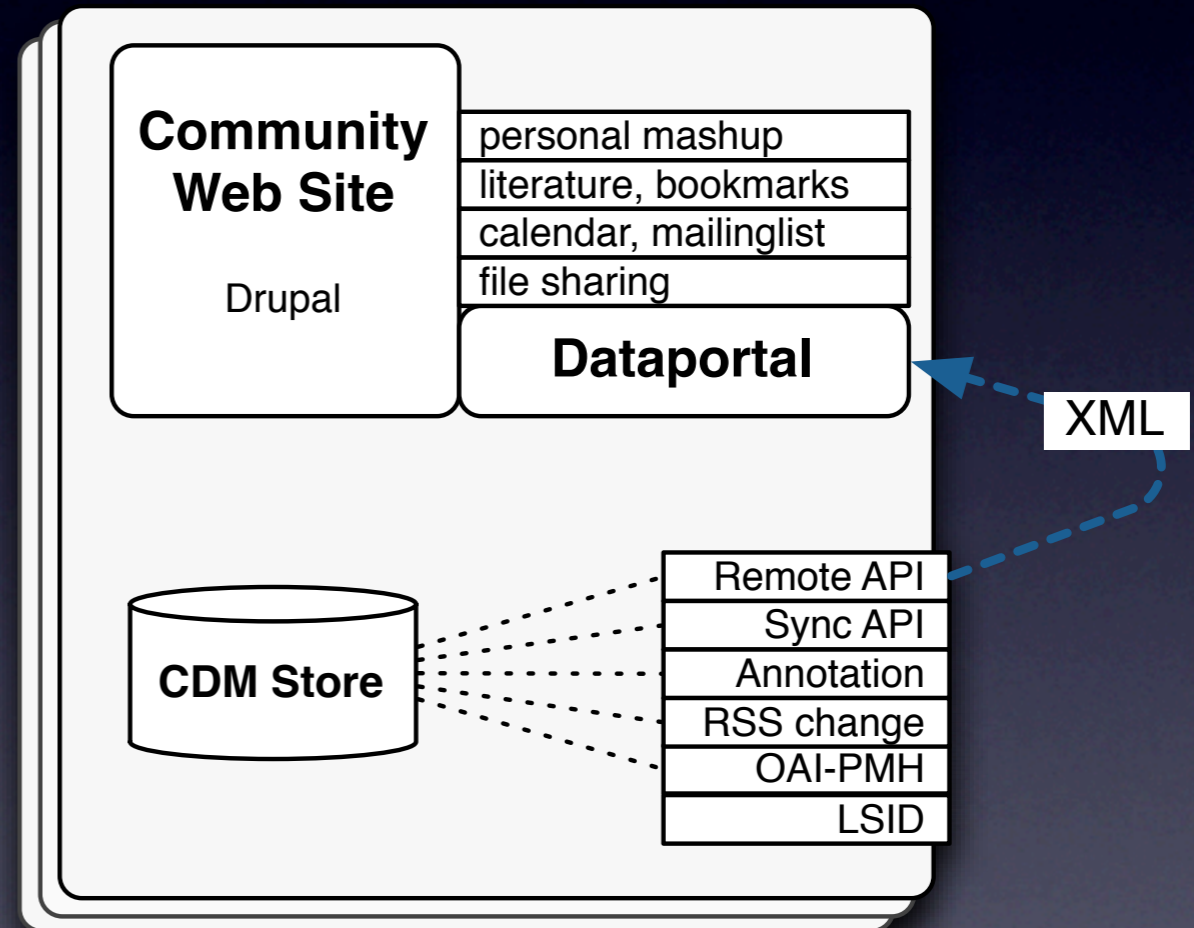## and the Common Data Model

Markus Döring
BGBM, FU-Berlin
October 2007

Community Sites

- Drupal Website
  - webpages, forum, blog
  - image+document hosting
- Dataportal
  - webpublication
- CDM Store
  - shared database
  - webservices

Community Web Site	personal mashup
	literature, bookmarks
	calendar, mailinglist
Drupal	file sharing
	Dataportal

XML

CDM Store
- Remote API
- Sync API
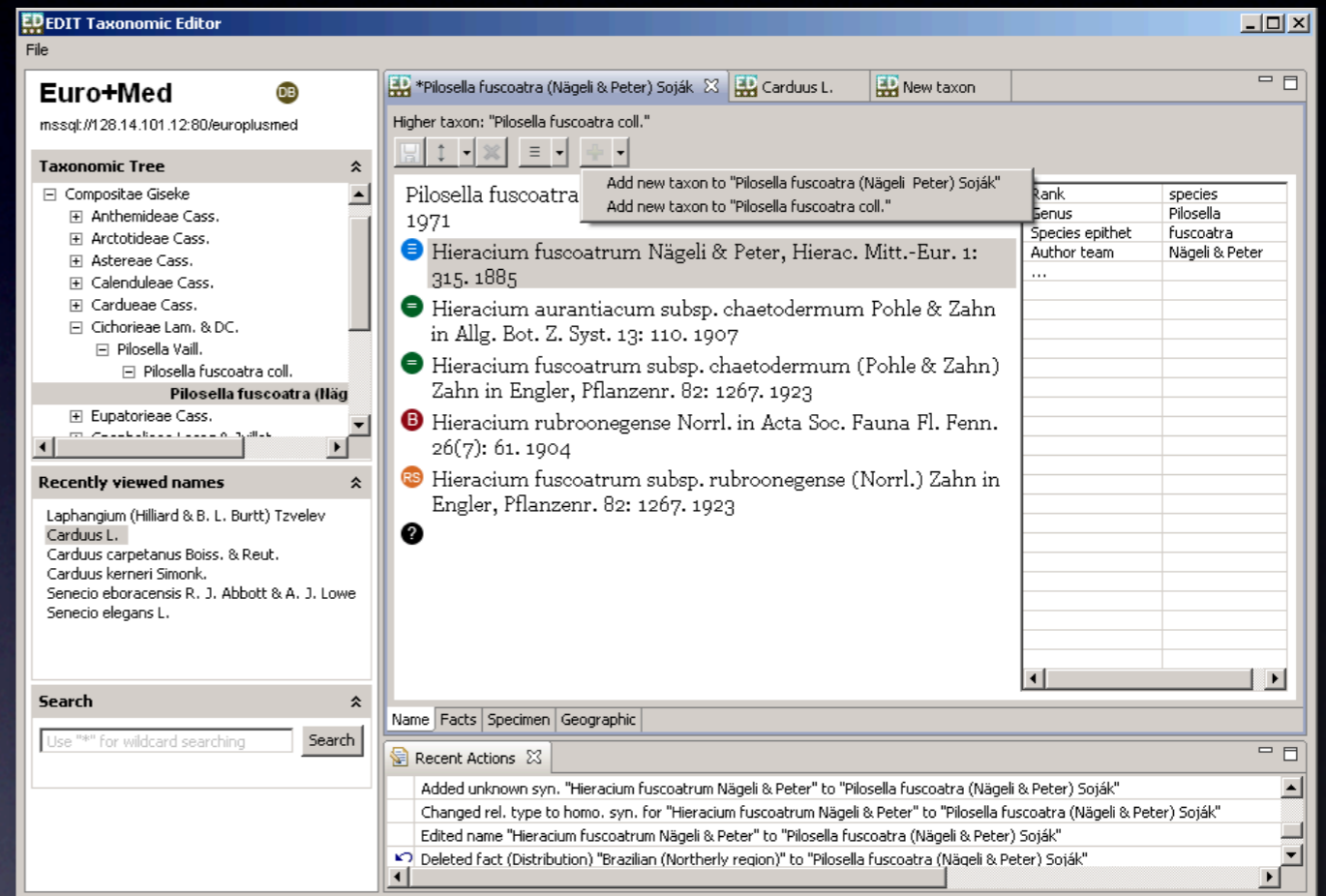- Annotation
- RSS change
- OAI-PMH
- LSID

Planned Dataportal

- Publication, no editing
  - concurrent concepts
- Drupal integration
- Uses CDM webservices
- Allows annotations
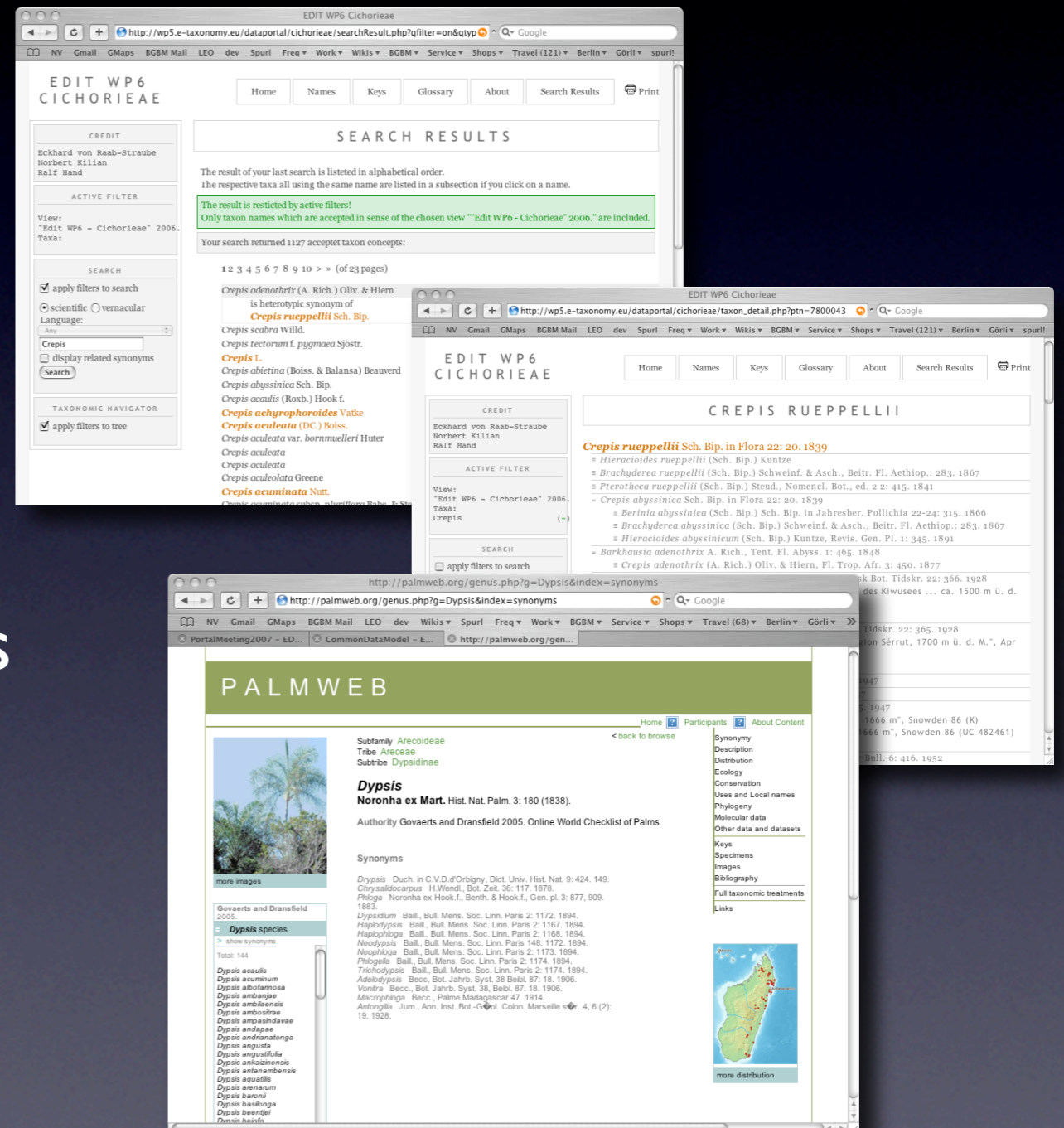
Taxonomic Editor

- Desktop App
  - offline work
  - Syncs w/ CDM Store
- Based on
  - Eclipse RCP
  - CDM library
  - embedded or LAN db
- Descriptions/Keys
  - initially via Xpert2 ?

Community CDM Store

- Central datastore for a community of users

- Uses CDM library w/ versioning

- Provides different APIs to interact with

  - REST read operations for portals
    - search/get for all identifiable classes. Specialised operations as required
  - RSS search feeds for subscriptions
  - Sync API, likely based on SyncML
  - OAI-PMH for aggregation
  - LSID resolver exposing objects in RDF as TDWG LSID voc
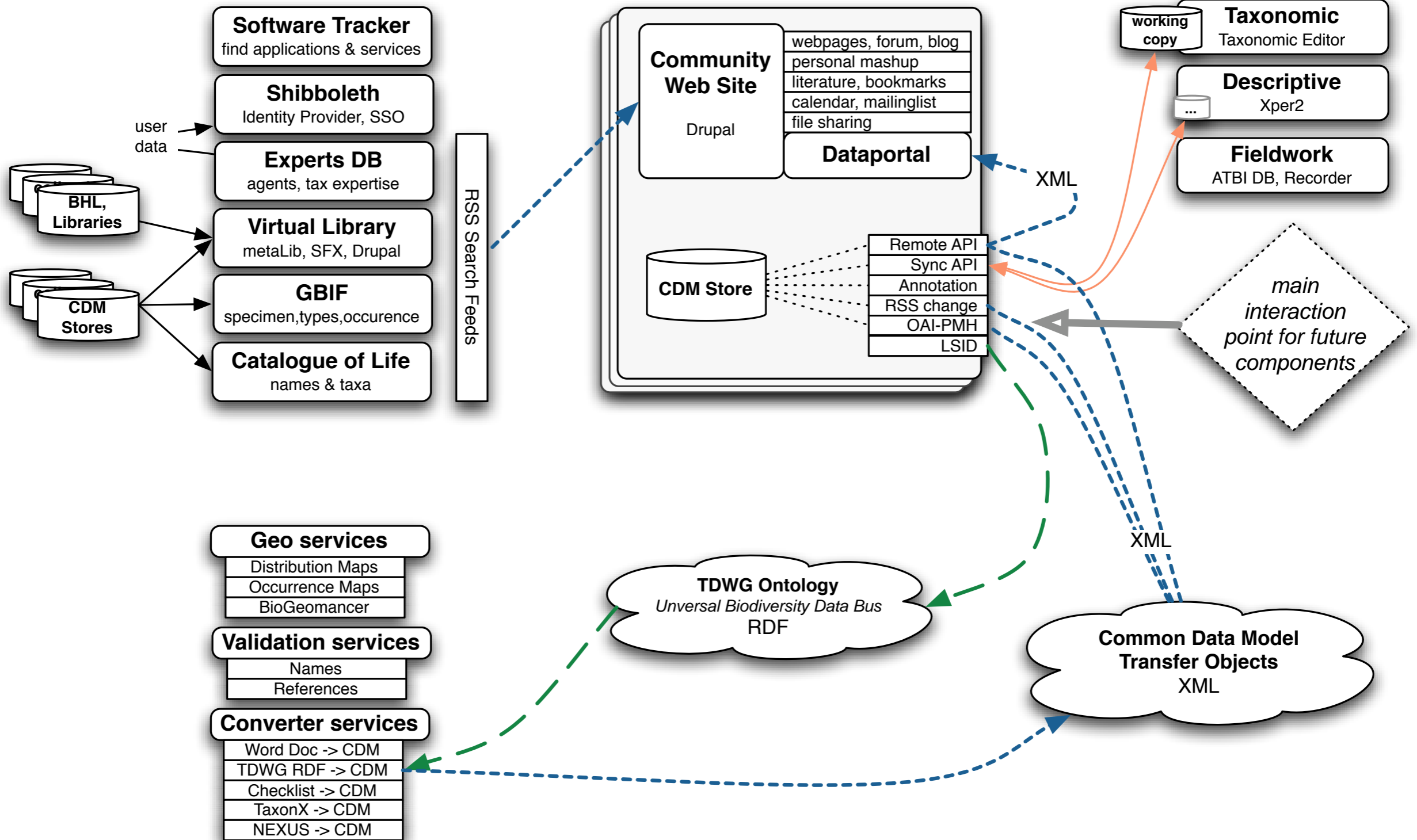
Other Components

- Converter:
  - TDWG, BibTex, NEXUS, Excel/CSV checklist, TaxonX, Word?
- Image server -> morphbank.org
- Geo services: online GIS app, distribution & occurrence map generation
- RSS search feeds: GBIF, ViTaL, Experts, Cat.of Life
- Virtual library (ViTaL)
- Experts DB (WP2)

EDIT Platform Components

Central Components

Software Tracker
find applications & services

Shibboleth
Identity Provider, SSO

user data →

Experts DB
agents, tax expertise

Virtual Library
metaLib, SFX, Drupal

GBIF
specimen, types, occurence

Catalogue of Life
names & taxa

BHL, Libraries

CDM Stores

RSS Search Feeds

Community Components

Community Web Site

Drupal

webpages, forum, blog
personal mashup
literature, bookmarks
calendar, mailinglist
file sharing

Dataportal

XML

CDM Store

Remote API
Sync API
Annotation
RSS change
OAI-PMH
LSID

Personal Components

working copy

Taxonomic
Taxonomic Editor

Descriptive
Xper2

...

Fieldwork
ATBI DB, Recorder

main interaction point for future components

Geo services

Distribution Maps
Occurrence Maps
BioGeomancer

Validation services

Names
References

Converter services

Word Doc -> CDM
TDWG RDF -> CDM
Checklist -> CDM
TaxonX -> CDM
NEXUS -> CDM

TDWG Ontology
Unversal Biodiversity Data Bus
RDF

Common Data Model Transfer Objects
XML

XML

Common Data Model (CDM)

- UML class model in Enterprise Architect

  - http://dev.e-taxonomy.eu/CommonDataModel

- Semantically aligns primarily with TDWG ontology.

  - but other standards like BibTex, SDD, DarwinCore considered

  - naming of classes & attributes

  - source for defined terms with identifiers. E.g. name ranks

- Roadmap

  - 4Q/2007 freeze model

  - 1Q/2008 release library and basic store

CDM Packages / Scope

- Common
- Reference
- Agent
- Name
- Taxon
- Location
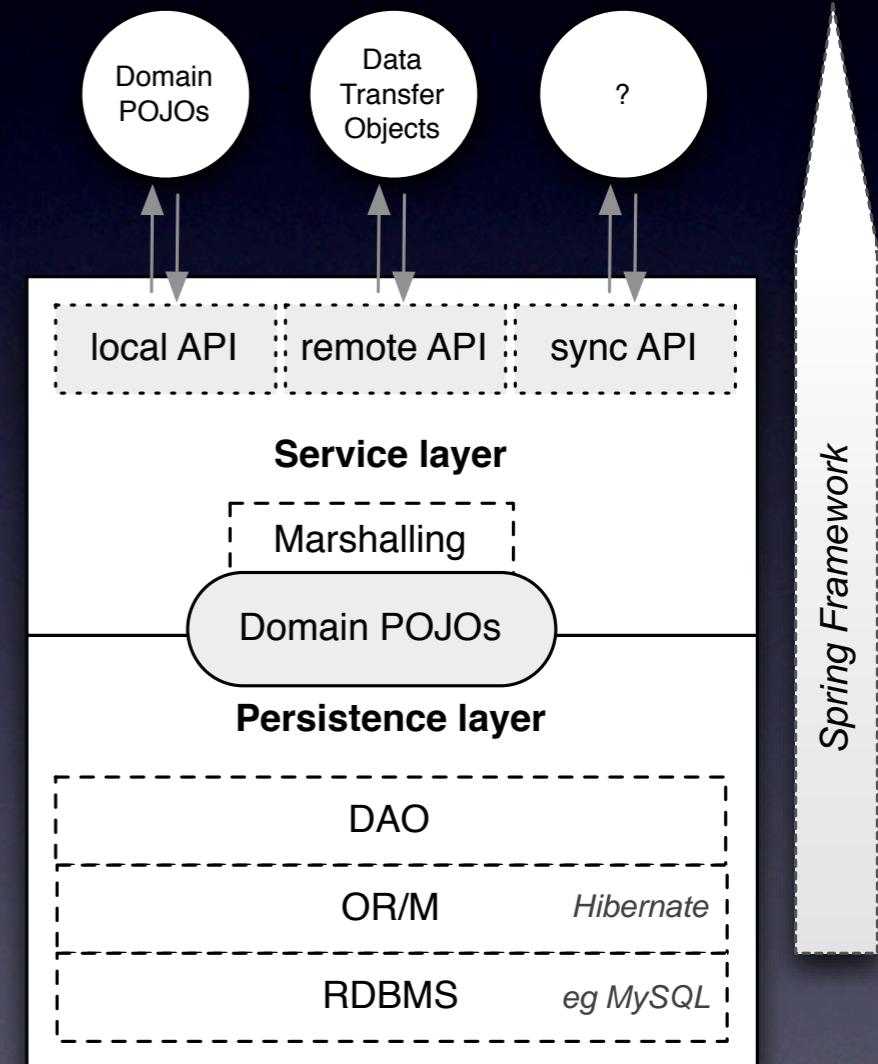- Occurrence
- Description
- Fact
- Molecular

class Model overview

common
- + AudioRecording
- + Copyright
- + *Defined TermBase*
- + *EntityInSourceBase*
- + Extension
- + ExtensionType
- + Figure
- + File
- + *IdentifiableEntity*
- + Image
- + Keyword
- + Language
- + Media
- + Movie
- + *ReferencedEntityBase*
- + Representation
- + Source
- + *VersionableEntity*
- + *IExtension*
- + *IFile*
- + *IMedia*
- + *IReferencedEntity*
- + *ITitle*

reference
- + Article
- + BibtexReference
- + Book
- + BookSection
- + CdDvd
- + ConferenceProceedings
- + Database
- + Generic
- + InProceedings
- + Journal
- + Map
- + Patent
- + PersonalCommunication
- + *PrintedUnitBase*
- + PrintSeries
- + Proceedings
- + *PublicationBase*
- + *ReferenceBase*
- + ReferenceInSource
- + Report
- + *SectionBase*
- + Thesis
- + WebPage
- + *INomenclaturalReference*
- + *IReference*

name
- + BacterialName
- + BotanicalName
- + CultivarPlantName
- + HybridRelationship
- + HybridRelationshipType
- + NameFactType
- + NameInSource
- + NameRelationship
- + NameRelationshipType
- + NameTypeDesignation
- + NomenclaturalStatus
- + NomenclaturalStatusType
- + NonViralName
- + Rank
- + SpecimenTypeDesignation
- + *TaxonNameBase*
- + TypeDesignationStatus
- + ViralName
- + ZoologicalName
- + *ITypeDesignation*

occurrence
- + Collection
- + Determination
- + Fossil
- + LivingIndividual
- + Observation
- + PreservationMethod
- + Specimen
- + TaxonOccurrence

location
- + Continent
- + NamedArea
- + NamedAreaLevel
- + NamedAreaType
- + Point
- + ReferenceSystem
- + WaterbodyOrCountry

molecular
- + Chromatogram
- + DnaSample
- + GeneBankAccession
- + Locus
- + PhylogeneticTree
- + Sequence
- + TissueSample

agent
- + Address
- + Contact
- + Institution
- + InstitutionalMembership
- + InstitutionType
- + Person
- + PersonInSource
- + Team
- + TeamInSource

taxon
- + ConceptRelationshipType
- + Synonym
- + SynonymRelationship
- + SynonymRelationshipType
- + Taxon
- + *TaxonBase*
- + TaxonCitation
- + TaxonInSource
- + TaxonRelationship

fact
- + Absent
- + CommonName
- + Distribution
- + *FactBase*
- + FactType
- + Introduced
- + NameFact
- + Native
- + OrganismInteraction
- + Present
- + TaxonFact

description
- + Character
- + CharacterTree
- + CodedDescription
- + DescriptiveConcept
- + IdentificationKey
- + NaturalLanguageDescription

EDIT
European Distributed Institute of Taxonomy

BGBM

CDM Vocabularies

- Extendible core terms

  - ranks, name & taxon relationship types, type designation

  - regions (TDWG), countries (ISO), languages (ISO)

  - occurrence status, collection & institution types

- Fixed terms as *classes*

  - nomenclatural code

  - reference type

- Still needed

  - "Fact" categories, species profile model categories: Economic Use, Morphology, Cytology, Biochemistry, ...

JUST IN CASE ...

CDM Java Library

- Persistent domain model
  - Plain Old Java Objects
  - Hibernate & Spring framework
  - XML marshalling
  - logic & methods
    - validation, matching, ...
- Flexible use
  - several APIs
  - chose backend, incl. embedded

Geo Services

- Map services

  - Occurrence points

  - Distribution per region presence/absence

- Online GIS app

  - multi species

  - targeted at ATBI sites