UNIVERSITY *of* CALIFORNIA, SAN DIEGO

MEDICAL CENTER MOORES CANCER CENTER

ERNE Grid Service

<http://erne.ucsd.edu:47210/wsrf/services/cagrid/UcsdErneData>

R. Hannes Niedner, M.D.

hniedner@ucsd.edu

caBIG Deployment Lead

Shared Resource for Biostatistics and Bioinformatics

PI: Karen S. Messer, Ph.D.

EDRN and caBIG

- Early Detection Research Network (EDRN) brings together dozens of institutions to help accelerate the translation of biomarker information into clinical applications and to evaluate new ways of testing cancer in its earliest stages and for cancer risk.

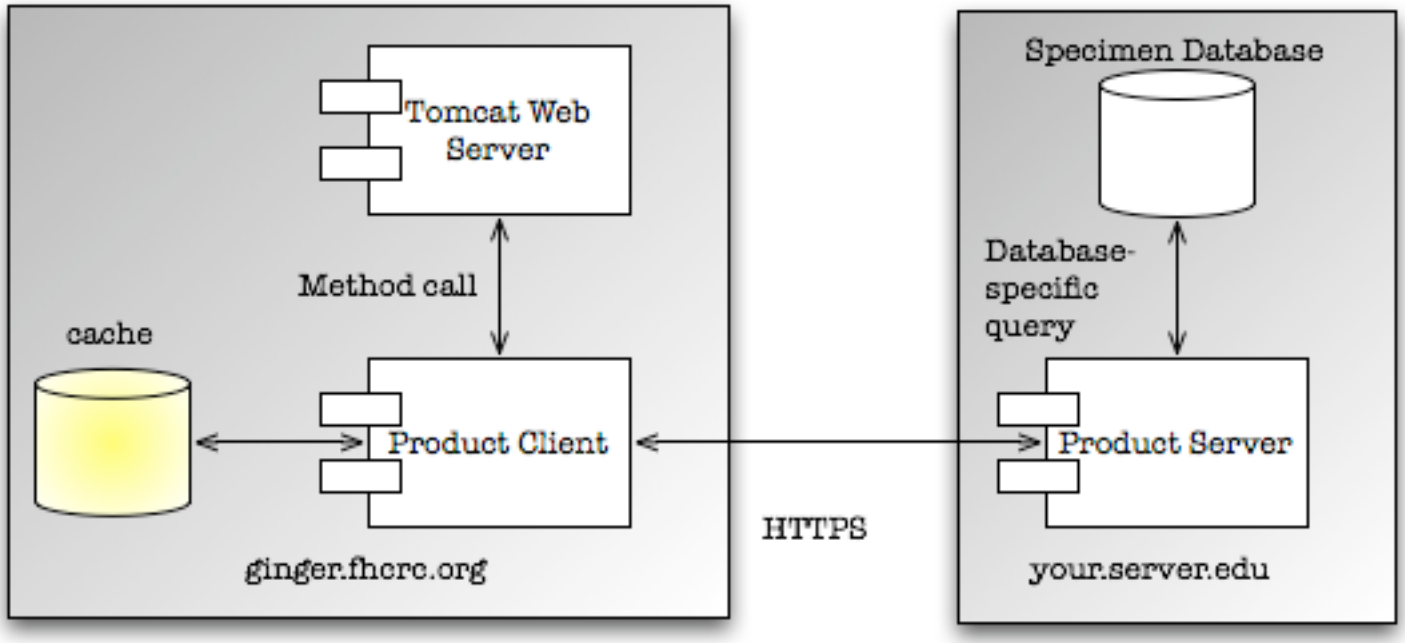
- The mission of caBIG® is to develop a truly collaborative information network that accelerates the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer, ultimately improving patient outcomes.

- Both are virtual organizations sponsored by the National Cancer Institute (NCI)

ERNE

The EDRN Resource Network Exchange (ERNE) is a virtual specimen bank that brings together existing disparate specimen databases into a unified whole, increasing the potential for scientific research.

Purpose: assist EDRN investigators with identifying potential collaborators who have specimens and associated epidemiological and clinical data of interest by allowing access to databases remotely via the EDRN website.



Goals

- Gather experience with the caCORE build process
- Share our ERNE data on caGRID
- Develop straight forward solution to get ERNE participants on caGRID
- Build a bridge between both networks

Motivation

- Shareable data a very hard to come by
- ERNE data are already shared
- Anonymized data, existing IRB approval
- Niche for caTissue (here at UCSD MCC)
  - Biobanking team happy with legacy application
  - Use caTissue grid service to share ERNE data
- Mismatch between ERNE and caTissue data model

EDRN CDEs v. caTissue Classes

SPECIMEN_AMOUNT-STORED_VALUE
SPECIMEN_AVAILABLE_CODE
SPECIMEN_COLLECTED_CODE
SPECIMEN_FINAL-STORE_CODE
SPECIMEN_STORED_CODE
SPECIMEN_SPUTUM_PRESERVATIVE_CODE
SPECIMEN_SPUTUM-PRESERVATIVE_OTHER_TEXT
SPECIMEN_TISSUE_DEGREE-INVASIVE_CODE
SPECIMEN_TISSUE_DEGREE-INVASIVE-TUMOR_CODE
SPECIMEN_TISSUE_ORGAN-SITE_CODE
SPECIMEN_TISSUE-ORGAN-SITE_OTHER_TEXT
SPECIMEN_AMOUNT-STORED_UNIT_CODE
SPEC_ID_NUM
BASELINE_CANCER_ICD9-CODE-OTHER_TEXT
BASELINE_CANCER-AGE-DIAGNOSIS_VALUE
BASELINE_CANCER-CONFIRMATION_CODE
BASELINE_CANCER-DIAGNOSIS_YEAR_TEXT
BASELINE_CANCER-ICD9-CODE
BASELINE_DATA-COLLECTED_DATE
BASELINE_DEMOGRAPHICS_RACE_CODE
BASELINE_DEMOGRAPHICS-BIRTH-YEAR_TEXT
BASELINE_DEMOGRAPHICS-GENDER_CODE
BASELINE_DEMOGRAPHICS-RACE_OTHER_TEXT
BASELINE_FAMILY_CANCER_CONFIRMATION_CODE
BASELINE_SMOKE-AVERAGE_DAY_VALUE
BASELINE_SMOKE-BEGIN-AGE_REGULAR_VALUE
BASELINE_SMOKE-QUIT-AGE_VALUE
BASELINE_SMOKE-REGULAR_1YEAR_CODE

FluidSpecimen
TissueSpecimen
CellSpecimen
MolecularSpecimen
SpecimenCharacteristics
SpecimenCollectionGroup
CollectionProtocolEvent
CollectionProtocolRegistration
Participant

anonymized

Unmapped (in caTissue Core)

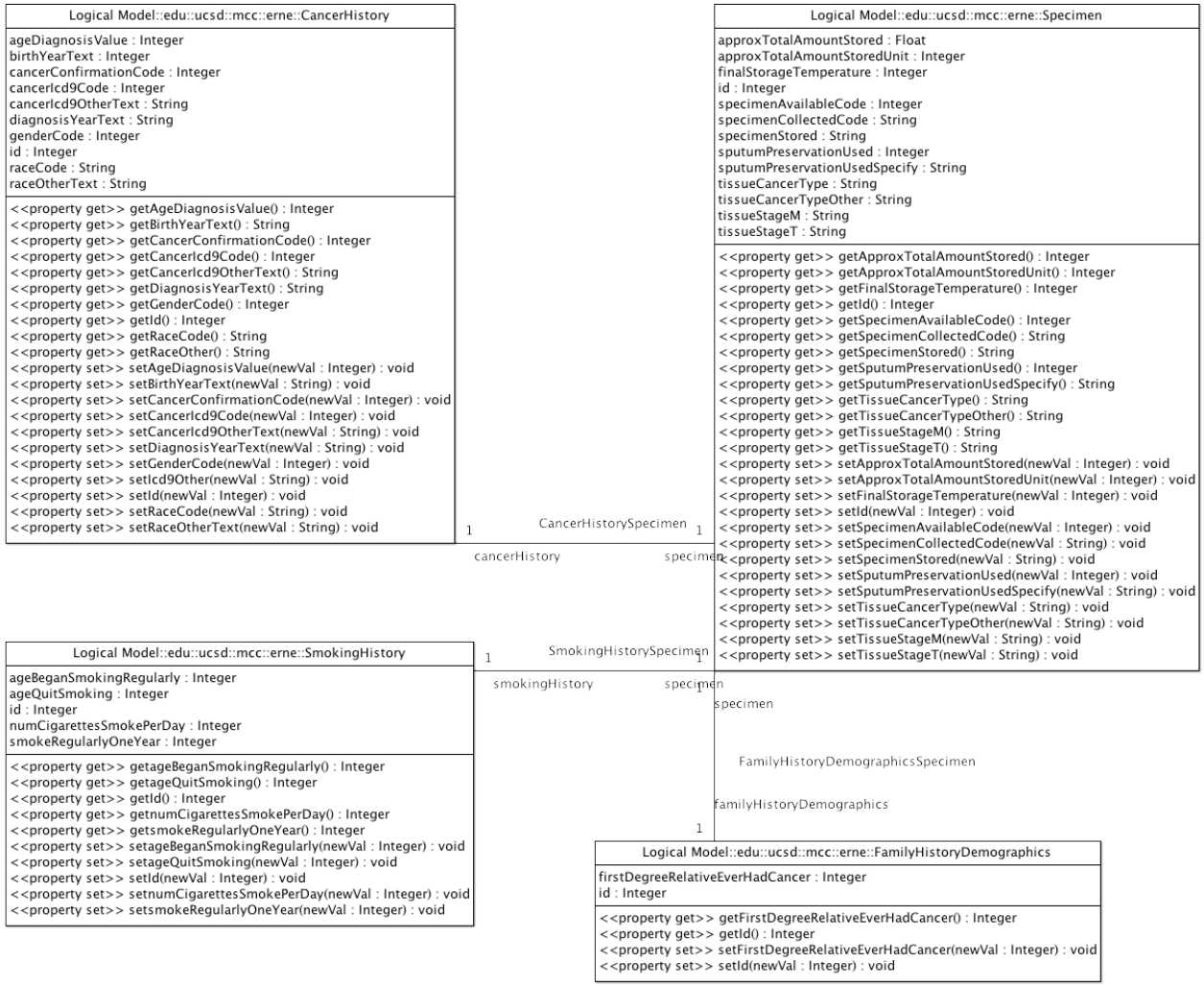
Process

- Develop UML model (ArgoUML or EA)
- Object-Relational mapping (caAdapter)
- Concept/CDE mapping (Semantic Integration Workbench)
- Code generation (caCORE SDK)
- Create Domain Model (ant xmiToDomainModel)
- Create Grid Service (Introduce)
- Specify Service Metadata (Introduce)
- Deploy grid service (Introduce)

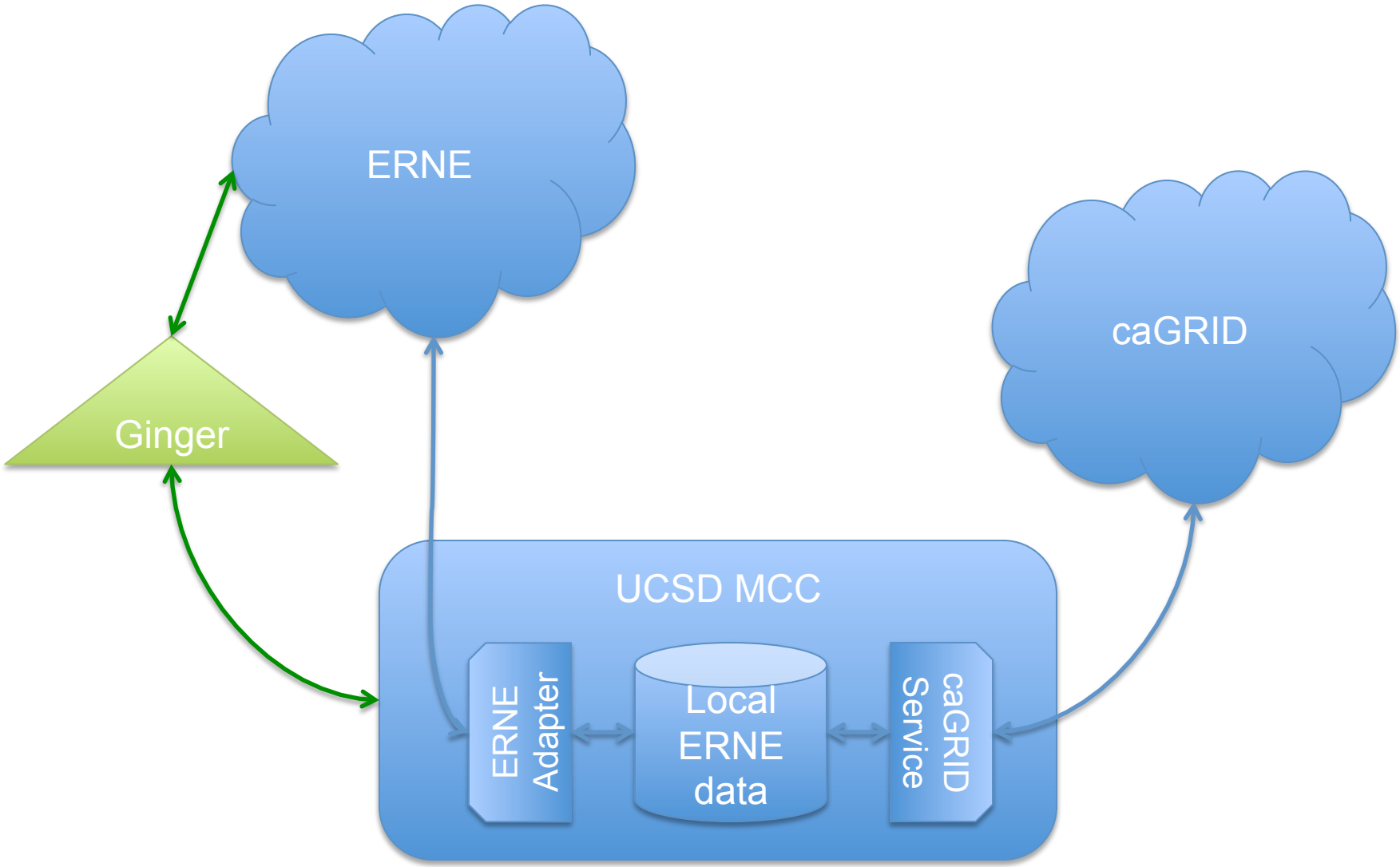
Some Problems

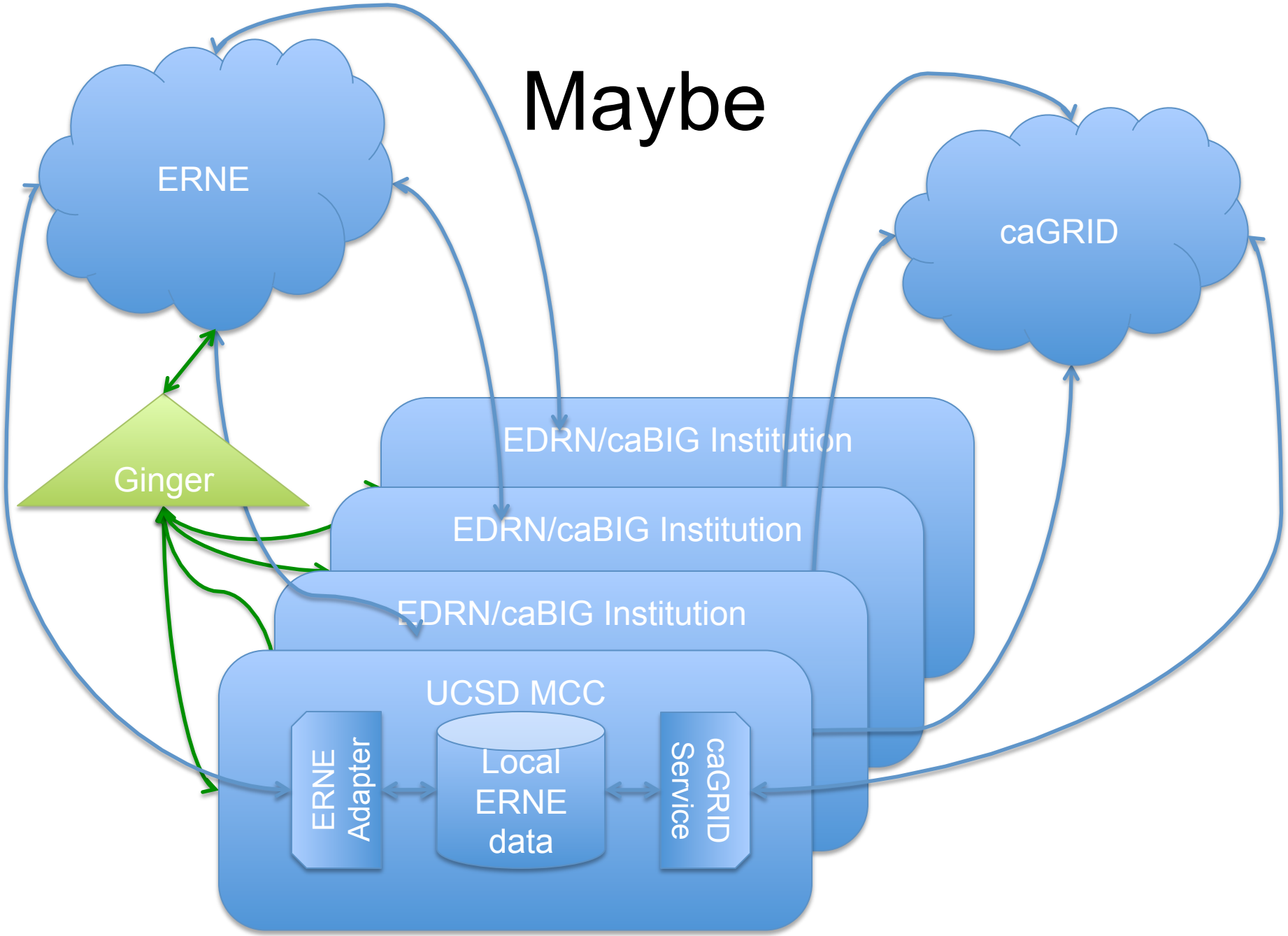
- Incomplete documentation of ERNE data model/CDEs
- No real-time communication with EDRN
- caCORE tools are still “beta” (impressive nonetheless)
- caCORE documentation doubly so
- Mismatch of local ERNE data representation with EDRN CDEs

Class Diagram

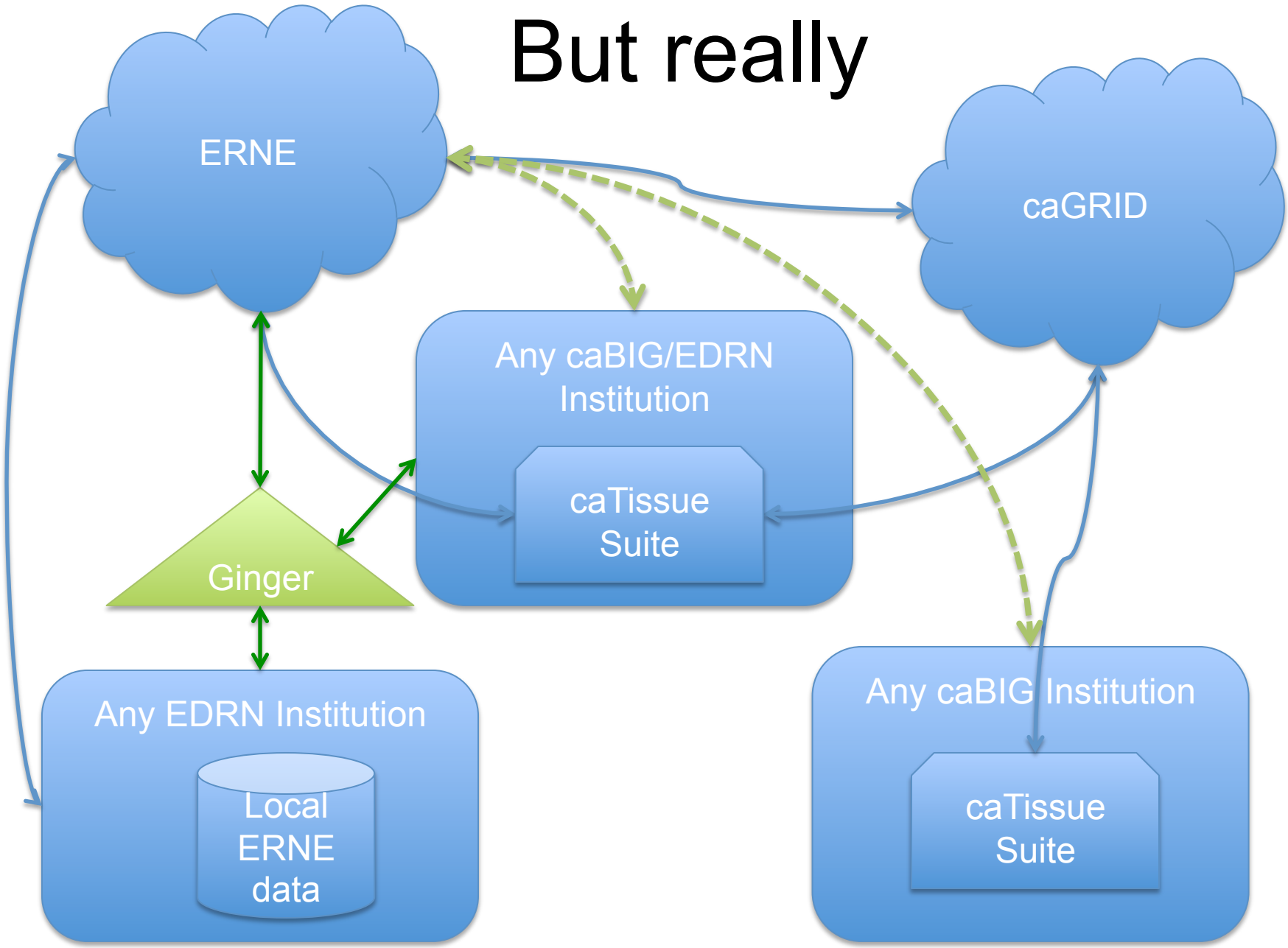


NOW


ERNE
caGRID
Ginger
UCSD MCC
ERNE Adapter
Local ERNE data
caGRID Service


Maybe
ERNE
caGRID
Ginger
EDRN/caBIG Institution
EDRN/caBIG Institution
EDRN/caBIG Institution
UCSD MCC
ERNE Adapter
Local ERNE data
caGRID Service

But really


ERNE
caGRID
Any caBIG/EDRN Institution
caTissue Suite
Ginger
Any EDRN Institution
Local ERNE data
Any caBIG Institution
caTissue Suite

Acknowledgements

- Karen S. Messer, Ph.D. – Director of Shared Resource for Biostatistics and Bioinformatics
- Richard Schwab, M.D., Biorepository PI
- Jerry Rowley – Biorepository DB Lead
- Tony Aung - Biorepository PA
- Sean Kelly, NASA JPL
- caGRID® and caCORE Knowledge Center Teams (especially Justin Permar, OSU)

My Questions

- What should be the “quasi standard model”?
- What will happen to the EDRN CDEs?
- Central or institutional grid service deployment?
- How many EDRN institutions are interested in sharing data on caGRID?
- What about caTissue Suite (data model, dynamic extension, federated queries)?

DEMO: Simple CQL

using the caGRID® Portal @ http://cagrid-portal.nci.nih.gov/web/guest/home

Retrieve all “specimen” records that are available.

```
<ns1:CQLQuery xmlns:ns1="http://CQL.caBIG/1/
  gov.nih.nci.cagrid.CQLQuery">
  <ns1:Target name="edu.ucsd.mcc.erne.Specimen">
  <ns1:Group logicRelation="AND">
   <ns1:Attribute name="specimenAvailableCode"
     predicate="EQUAL_TO" value="1"/>
  </ns1:Group> </ns1:Target>
</ns1:CQLQuery>
```