The Chemical Analysis Metadata Platform (ChAMP)
(http://champ-project.org)

Stuart J. Chalk, Department of Chemistry, University of North Florida
Antony Williams and Valery Tkachenko, RSC Cheminformatics
schalk@unf.edu
Overview

- Initial Idea
- Motivation
- Why a Platform?
- Pieces of the Puzzle
- Existing Resources
- What are the Most Important Metadata?
- Minimum Information About a Chemical Analysis?
- Ontology Development
- Example Application
- Future Developments
- Conclusion
Initial Idea

* Develop a set of metadata items for representation/annotation of chemical analysis information

* Are there important characteristics (metadata) about analysis methodologies that, if captured, would add value to a resource?

* Must be easy to implement

* Must be useful across multiple disciplines
Motivation

- How to facilitate aggregation/searching of CA information?
  - Knowledge in existing literature
  - Annotation of research in future publications
  - Annotation of (potentially useful) unpublished/self published work
  - Annotation of data captured in ELN’s
- Need tool to annotate data in digital repositories
  - Provide users with uniform (but flexible) mechanism to categorize data they contribute
  - Help researchers articulate data management plans in grants
- Complement/extend existing activities

- The haystack is so big – we need to make it easy to visualize the needle by accurate annotation of available methodologies
RSC Data Repository

User interface layer (examples)

- ChemSpider 2.0
- Paid 3rd party integrations (various platforms - SharePoint, Google, etc)
- Electronic Laboratory Notebook

Analytical Laboratory application

Chemical Inventory application

User interface widgets layer

- Compounds Widgets
- Reactions Widgets
- Spectra Widgets
- Crystals Widgets
- Documents Widgets

Data access layer

- Compounds API
- Reactions API
- Spectra API
- Crystals API
- Documents API

Data layer

- Compounds
- Reactions
- Spectra
- Crystals
- Documents
Look at the posts for analytical method help on Linked-In

· ‘I need an ICP-MS application note about direct determination of sulfur and phosphate in microwave digested plant material and soil without using external oxygen as a reaction gas.’ (ICP-OES and ICP-MS)

· ‘I want to validate a method of detecting As in glass vials with the aid of atomic absorption and air-acetylene flame’. (Analytical Method Validation)

· ‘Does anyone know another method for determining total iron and copper in water other than calorimeter and wet chemistry?’ (Analytical Chemistry)

· ‘Anyone with knowledge in electrochemical detection of Homovanillic acid in urine samples?’ (Analytical Chemistry)
Why a Platform (Toolkit)?

* Develop it to be as broadly applicable as possible
* Chemical analysis is a not tangible like a spectrum
* Users have domain specific needs/goals
* Users has a favorite/required format to store information
  * SQL Relational Database, No-SQL, Excel Spreadsheet
  * XML, YAML, JSON or JSON-LD
* Allows use in different ways – facilitates usage
  * Build a new data standard using ChAMP
  * Annotate an existing data standard
* ChAMP should define the types of metadata and general organization of the information, not the format it is stored in (this is like MIAME [1])

First Thoughts

* Covers metadata for a chemical analysis methodology not raw analytical instrument data
* Use existing technology/standards where-ever possible
* Nothing is required – some things highly recommended
* Can use all of specification, some parts, or only one piece
* Useful for both method development and application
* Platform scope should be as wide as possible

* What information is most important?
* How do we get community involvement/buy-in?
Pieces of the Puzzle

- Description of important CA metadata
- Taxonomy of CA metadata
- Ontology of chemical analysis terms
  - Broad terms initially
  - Development of technique specific terms/concepts later
- Controlled vocabularies for specific metadata items
- Definitions of required metadata (in context)
- Naming and design rules
Existing Resources

* Ontologies
  * Chemical Methods Ontology (CMO) [2]
  * SemanticScience CHEMINF Ontology [3]
  * Chemical Entities of Biological Interest (ChEBI) [4]
  * Basic Formal Ontology [5]

Existing Resources

* Controlled Vocabularies/Taxonomies
  * MESH [6]
  * LCSH [7]
  * CAS Subject Headings [8]
  * IUPAC Orange Book [9]
  * IUPAC Gold Book [10]
  * ... do they address how to organize the metadata?

Existing Resources

Other

* JCAMP-DX [11]
* Analytical Information Markup Language (AnIML) [12]
* Units Markup Language (UnitsML) [13]
* NASA Quantities, Units, Dimensions and Data Types [14]
* Electronic Laboratory Notebook Manifest (elnItemManifest) [15]

What are the Most Important Metadata?

* Depends on who you talk to...

* Platform should describe (as completely as possible) the types of metadata important in analysis...

* ... but leave the description of what’s important to the users

* Standards for different industries, with different requirements, could be developed based on the platform
Minimum Information About a Chemical Analysis?

- MIAChA (my-ache-a?)

- Can the community agree on a minimum set of metadata items needed to annotate an analysis?

- Must be for a more specific area of analysis
  - MIASA – Spectrochemical Analysis
  - MIACA – Chromatographic Analysis
  - MIAEA – Electrochemical Analysis
  - MIATA – Thermal Analysis
Chemical Analysis Ontology

- An ontology to represent the concepts in the discipline of chemical analysis AND the metadata and data structures important to the area
- Borrows heavily from
  - Chemical Methods Ontology
  - Chemical Information Ontology
  - Chemical Entities of Biological Interest Ontology
  - Basic Formal Ontology
  - Unit of Measure Ontology
Chemical Analysis Ontology
Chemical Analysis Ontology

- Analysis metric datum
  - Coefficient of determination
  - Dynamic range
  - Limit of detection
  - Limit of linearity
  - Limit of quantitation
  - Linear dynamic range
  - Repeatability
  - Ruggedness
  - Sample throughput
  - Sensitivity
  - Signal to noise ratio
  - Specificity
  - Calculated result data item
  - Calculated result datum
- Chemical descriptor
- Chemometric datum
  - Chi square test
  - F-test
  - Non-parametric test
  - One-way ANOVA
  - Paired t-test
  - Student’s t-test
  - Two-way ANOVA
  - Equipment specification datum
- Instrument descriptor
- Instrument setting datum
- Measurement data item
  - Spectrum
  - Time course
    - Chromatogram
    - Flagram
    - Kinetics trace
- Concept
  - Analyte
    - Analyte class
  - Analytical technique
    - Instrumental technique
    - Remote technique
    - Sensor technique
    - Wet chemical technique
  - Analyzed form
  - Application area
- Chemical analysis
  - Functional group test
  - Property measurement
  - Qualitative analysis
  - Quantitative analysis
  - Structure elucidation
  - Deployment location
  - Figure of merit
- Interference
  - Interference (Different mechanism)
  - Interference (Similar mechanism)
- Matrix
- Mixture
- Property
  - Bulk property
  - Chemical property
Chemical Analysis Ontology

- 'Material entity'
- 'Analytical entity'
  - 'Analytical instrument'
  - 'Analytical instrument accessory'
  - 'Analytical instrument component'
  - 'Chemical sensor'
  - 'Hyphenated analytical instrument'
  - 'Portable analytical instrument'
  - 'Remote analytical instrument'
  - 'Wet chemical analysis apparatus'
- 'Chemical entity'
- 'Field sampling entity'
  - 'Dip net'
  - 'Dredge'
  - 'Grab sampler'
  - 'Preservative'
  - 'Sample container'
- 'Laboratory entity'
  - 'Analytical glassware'
  - 'Analytical instrument'
  - 'Equipment'
- 'Materials'
  - 'Calibration standard'
  - 'Primary standard'
  - 'Reagent'
  - 'Sample'
  - 'Solution'
  - 'Specimen'
  - 'Non-analytical glassware'
  - 'Non-analytical instrument'

- 'Role'
  - 'Analysis role'
    - 'Analyte role'
    - 'Calibrant role'
    - 'Interferent role'
    - 'Matrix role'
    - 'Primary standard role'
    - 'Secondary standard role'
    - 'Spike role'
    - 'Standard role'
  - 'Application role'
    - 'Environmental role'
    - 'Medical role'
    - 'Pharmaceutical role'
    - 'Biological role'
  - 'Chemical role'
  - 'Chemistry role'
  - 'Instrument role'
    - 'Detection role'
    - 'Sampling role'
    - 'Separation role'
  - 'Person role'
    - 'Analyst role'
    - 'Group leader role'
    - 'Laboratory manager role'
    - 'Principal investigator role'
  - 'Reaction role'
Example Application

- Summary information for a journal article
- Implementing ChAMP in XML

- ChAMP XML Schema
- Journal Article Metadata Specification Schema
- Instance file (XML file for one journal article)
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
    xmlns="http://champ-project.org/journal"
    xmlns:champ="http://champ-project.org/champ"
    xmlns:dcterms="http://purl.org/dc/terms/
    elementFormDefault="qualified" attributeFormDefault="unqualified"
    targetNamespace="http://champ-project.org/journal" version="1.0" xml:lang="en">
    <xs:import namespace="http://champ-project.org/champ" schemaLocation="champ.xsd"/>
    <xs:element name="overview" substitutionGroup="champ:description"/>
    <xs:element name="article" type="articleType"/>
    <xs:complexType name="articleType">
        <xs:sequence>
            <xs:element ref="overview" maxOccurs="1"/>
            <xs:element ref="champ:contact" maxOccurs="unbounded"/>
            <xs:element ref="champ:analyte" maxOccurs="unbounded"/>
            <xs:element ref="champ:matrix" maxOccurs="unbounded"/>
            <xs:element ref="champ:samplingConditions" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="champ:instrument" maxOccurs="unbounded"/>
            <xs:element ref="champ:metric" minOccurs="0" maxOccurs="unbounded"/>
            <xs:element ref="champ:concept" minOccurs="0" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:schema>
<article xmlns="http://champ-project.org/journal"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:champ="http://champ-project.org/champ"
  xmlns:dc:"http://purl.org/dc/elements/1.1/"
  xsi:schemaLocation="http://champ-project.org/journal champ_article.xsd">
  <overview champ:id="&CAO;CA0_000002">
    <dcterms:title>Plasticized Poly(vinyl chloride)-Based Photonic Crystal for Ion Sensing</dcterms:title>
    <champ:focus>Inorganic materials from ion analysis</champ:focus>
    <dcterms:bibliographicCitation>Anal. Chem., 2014, 86 (24), pp 11986-11991 DOI:10.1021/ac503447m</dcterms:bibliographicCitation>
  </overview>
  <champ:contact>
    <champ:person champ:id="http://xmlns.com/foaf/0.1/Person">Tatsuro Endo</champ:person>
    <champ:address>Department of Applied Chemistry, Osaka Prefecture University, 1-1 Gakuencho, Naka-ku, Sakai, Osaka 599-8531, Jap</champ:address>
    <champ:email>endo@chem.osakafu-u.ac.jp</champ:email>
    <champ:phone>+81-72-254-9284</champ:phone>
    <champ:role>Corresponding Author</champ:role>
  </champ:contact>
  <champ:analyte champ:id="&CAO;CA0_000004">
    <champ:substance champ:id="&CI;CHEMINF_000266">
      <champ:inchiString>"InChI=1S/K+1"
      <champ:inchiKey>"NPYPAHNBDSXSS-UHFFFAOYSA-N"
      <champ:substanceName>Potassium ion</champ:substanceName>
    </champ:substance>
  </champ:analyte>
  <champ:matrix champ:id="&CHMD;CHM0_0002743">Buffer Solution</champ:matrix>
  <champ:samplingConditions champ:encoding="json">"temperature"=>23.7'C,'pressure'=>1 atm}</champ:samplingConditions>
  <champ:instrument>Polymor-based Optical Sensor</champ:instrument>
  <champ:instrument>Visible spectroscopy</champ:instrument>
  <champ:concept>
    <champ:term champ:id="&ORO;ORCS_00000058">sensitivity</champ:term>
  </champ:concept>
  <champ:scope champ:general="champ:scope">
    <champ:source champ:id="doi:10.0001/fakedoi">ChAMP Concept Vocabulary</champ:source>
  </champ:concept>
</article>
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
    xmlns="http://champ-project.org/journal"
    xmlns:champ="http://champ-project.org/champ"
    xmlns:dcterms="http://purl.org/dc/terms/"
    elementFormDefault="qualified" attributeFormDefault="unqualified"
    targetNamespace="http://champ-project.org/journal" version="1.0" xml:lang="en">

    <xs:import namespace="http://champ-project.org/champ" schemaLocation="champ.xsd"/>
    <xs:element name="summary" substitutionGroup="dcterms:abstract"/>
    <xs:element name="qualityControl" substitutionGroup="champ:procedure"/>

    <xs:element name="stdMethod" type="methodType"/>

    <xs:complexType name="methodType">
        <xs:sequence>
            <xs:element ref="champ:analyte" maxOccurs="unbounded"/>
            <xs:element ref="champ:scope" maxOccurs="1"/>
            <xs:element ref="champ:applicationArea" maxOccurs="1"/>
            <xs:element ref="summary" maxOccurs="1"/>
            <xs:element ref="champ:interferences" maxOccurs="unbounded"/>
            <xs:element ref="champ:instrument" maxOccurs="unbounded"/>
            <xs:element ref="champ:reagent" maxOccurs="unbounded"/>
            <xs:element ref="champ:reagentSolution" maxOccurs="unbounded"/>
            <xs:element ref="champ:samplingProtocol" maxOccurs="1"/>
            <xs:element ref="champ:storageConditions" maxOccurs="unbounded"/>
            <xs:element ref="champ:analysisTimeframe" maxOccurs="1"/>
            <xs:element ref="champ:procedure" maxOccurs="1"/>
            <xs:element ref="qualityControl" maxOccurs="1"/>
            <xs:element ref="dcterms:bibliographicCitation" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
</xs:schema>
Future Developments

* Publish version 1 of platform (with best practices)
* General Concept Vocabulary for Chemical Analysis
* Concept Vocabularies for Specific Techniques
  * Repurpose any existing vocabularies (with permission)
  * Convert/integrate IUPAC ‘terminology’ publications
* Provide example documents in different formats
* Additional example applications
  * Partner with groups in different areas
Conclusion

* The ‘platform’ approach will make it easier for scientists to
  * Develop new standards for representing chemical analysis information
  * Integrate semantic annotation into exiting standards
* It will enhance basic searching (through standardization and vocabularies)
* It will allow semantic searching
* It will provide efficient annotation of large amounts of curated data that is not from traditional publishing
* Fits with the mission of the Research Data Alliance [16]

Questions?

* schalk@unf.edu
* Phone: 904-210-5311
* Skype: stuartchalk
* LinkedIn/Slidehare: https://www.linkedin.com/in/stuchalk
* ORCID: http://orcid.org/0000-0002-0703-7776