G-SRS2.0 and Beyond

Tyler Peryea

NATIONAL CENTER FOR ADVANCING TRANSLATIONAL SCIENCES
NATIONAL INSTITUTES OF HEALTH
11 OCTOBER 2017

NCATS

NIH National Center for Advancing Translational Sciences

Outline

- GSRSv1.3 to GSRSv2.0
  - Technical Challenges
  - Updated Features / Changes for public GSRS
  - Updated Features for FDA GSRS
- API consumers and other tools
- Next Steps

GSRS1.3 to GSRS2.0

- January: First full production release
  - **GSRS 1.1** used in parallel with legacy system
- March: Second full production release
  - **GSRS 1.2** fully replaces legacy system
- July: Minor feature update release to FDA
  - **GSRS 1.3** new integrations, support for more legacy tools

- Future:
  - **GSRS 2.0** in pre-production at FDA
  - Slated for production later this month

GSRS in production at FDA

- **ISO IDMP 11238-based substance records with**
  - ~100,000 expert-validated substance records
  - ~15,000 concepts
  - ~70,000 pending non-validated substance records
  - ~4,000 active users
  - Robust set of synonyms, codes, and classifications
  - Substance relationships including:
    - Metabolites
    - Impurities
    - Active Moiety relationships
    - Chemical constituents found in plant and animal material

GSRS in production at FDA

- **Base Features:**
  - JSON model for substances
  - REST API for programmatically searching, browsing, updating and validating substance records
  - Structure and sequence search support
  - Validation rules for detecting substance definitions that are sufficiently similar
  - Facet on the UI and the REST API
  - Support for many common export formats
  - Full audit history for every edit
  - CV and pick-list sets for relevant terminology
  - Enhanced structure-based CV for select substance classes

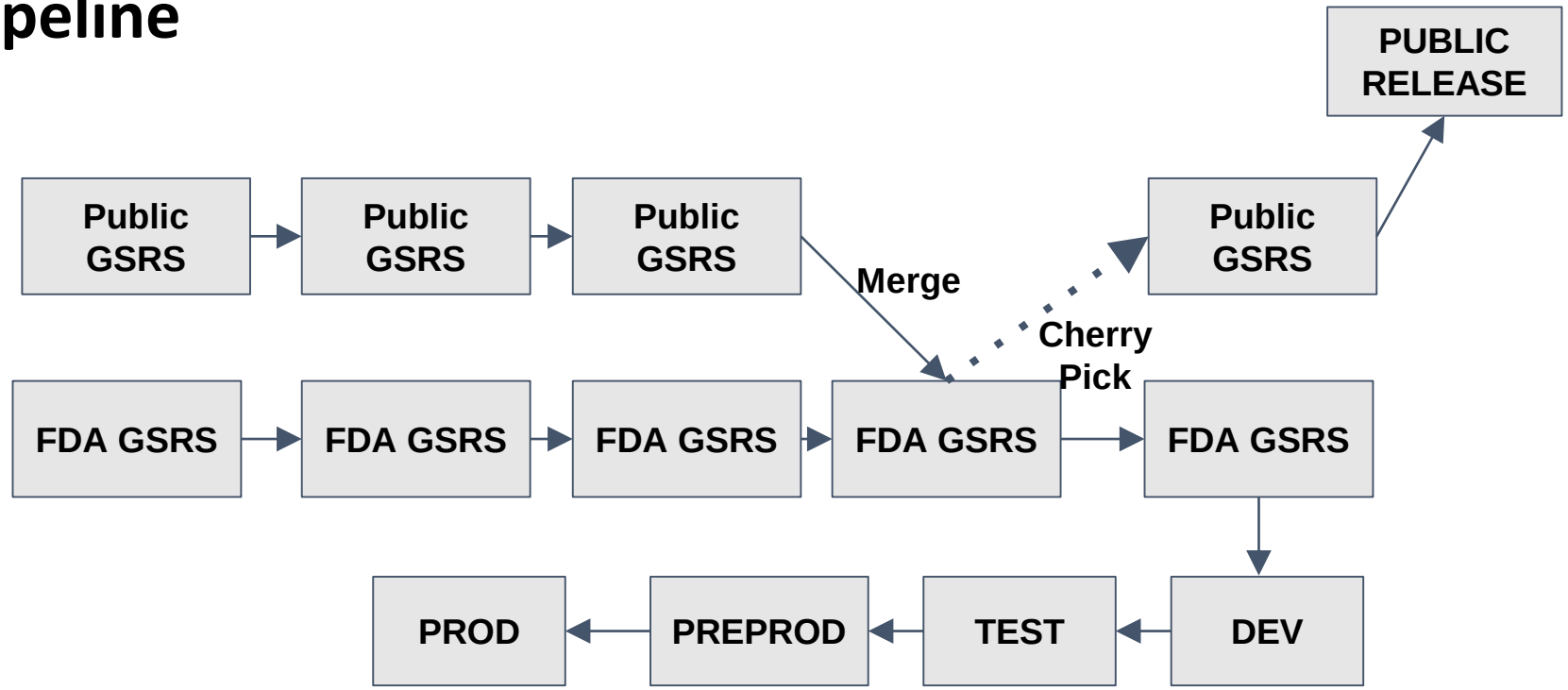
FDA GSRS v1.X Desired Enhancements

- **Enhancements and expansion of Facets**
- **More contextual, scientific and medical information integrated**
- **Simplified forms for common simple tasks (e.g. adding synonyms)**
- **Group 2, 3 and 4 Specified Substance support**
- **Automated reports / exports for specific data pipelines**
- **Modified validation rules**
- **Structure-based searching for mixtures**

Technical Challenges

•Old Development Pipeline

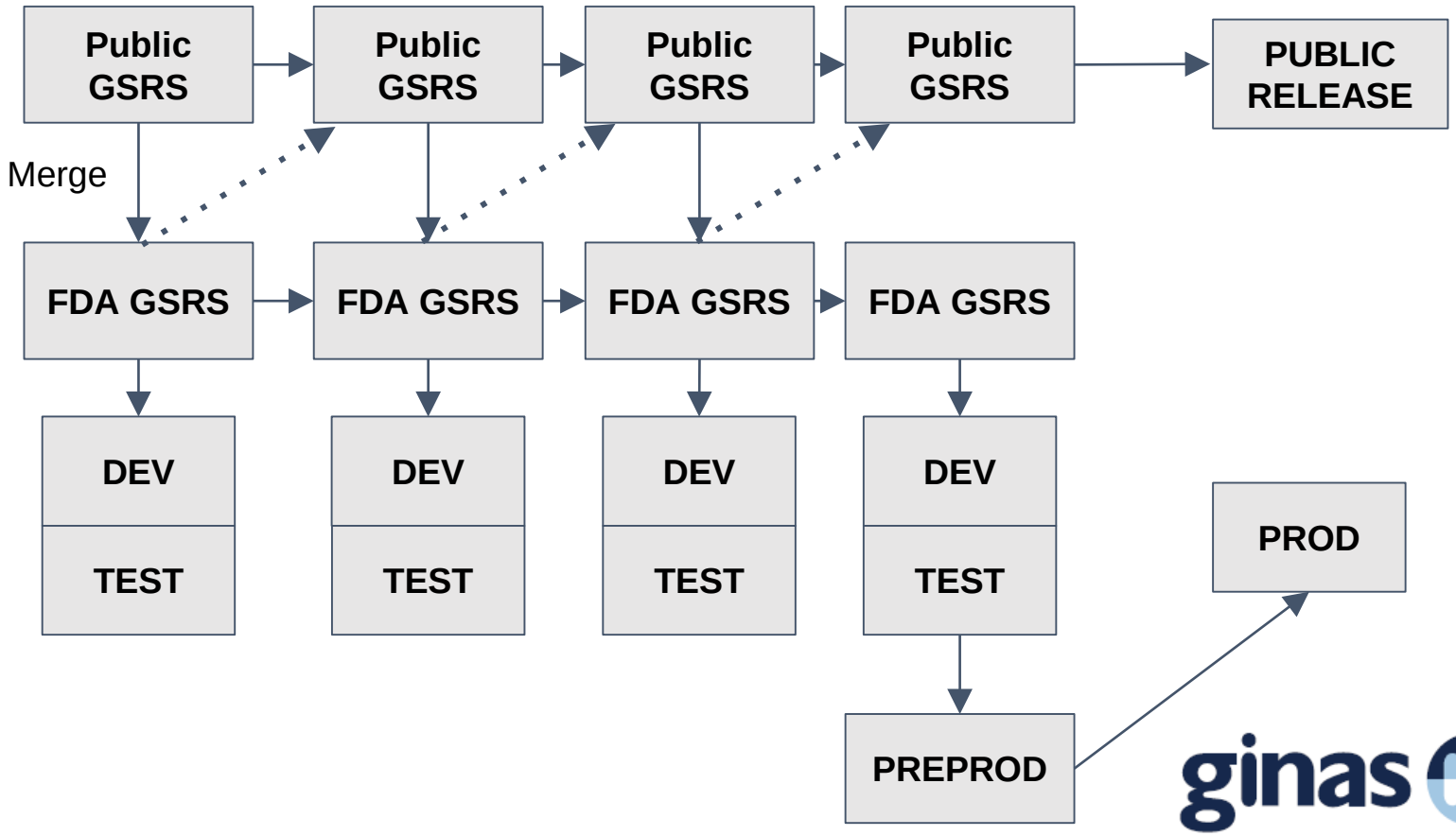
- **Pros:**
  - Insulates public GSRS from code *very* specific to FDA
- **Cons:**
  - When merges occur, they consume a lot of time
  - Pipeline to user-testable build fairly slow (about ***every 3 weeks***)



Technical Challenges

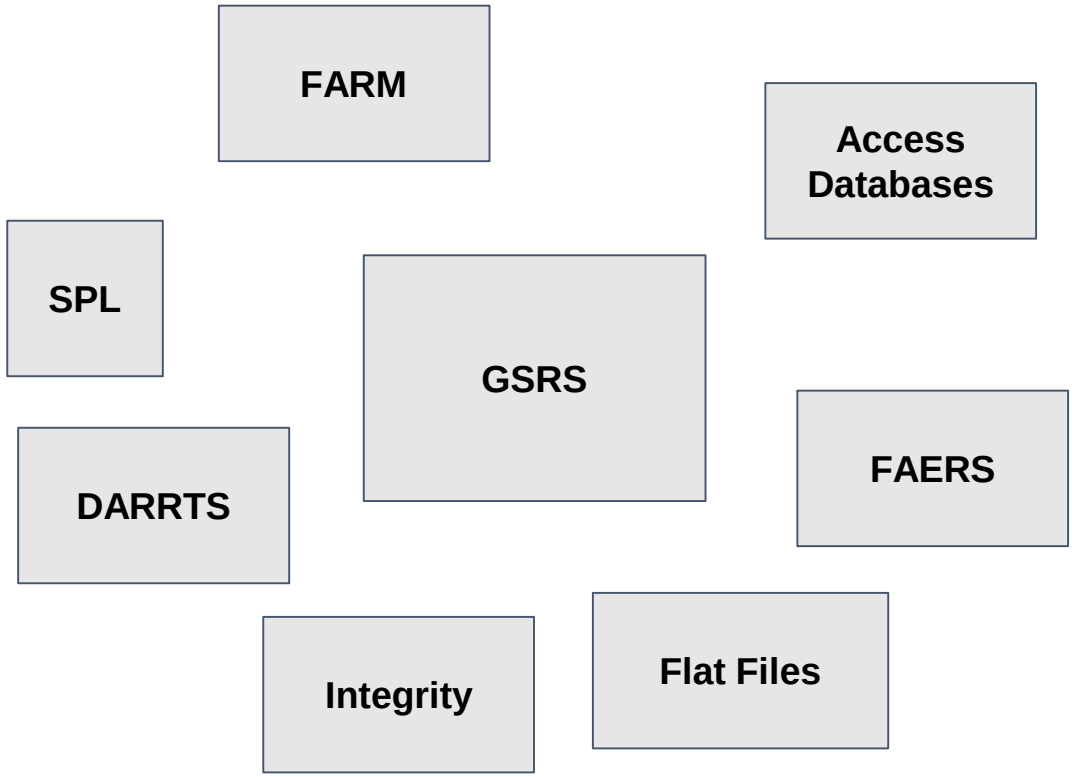
•*New* Development Pipeline

- **Pros:**
  - Merge conflicts are rarely an issue
  - Rapid deployment to user-facing environment (typically ***once per day***)
  - Cherry picks are easier
- **Semi-Con:**
  - Public GSRS development closely tied to FDA GSRS development (which is also good)



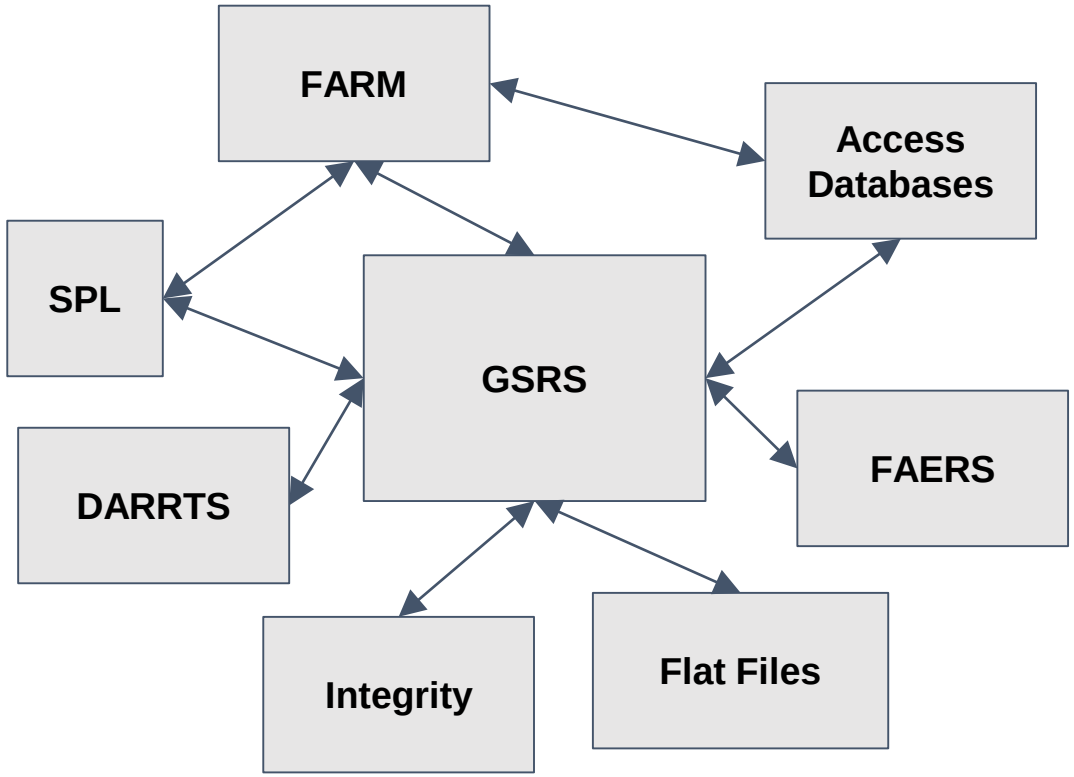
Technical Challenges (cont)

- The main purpose of GSRS is to represent substances faithfully and uniquely



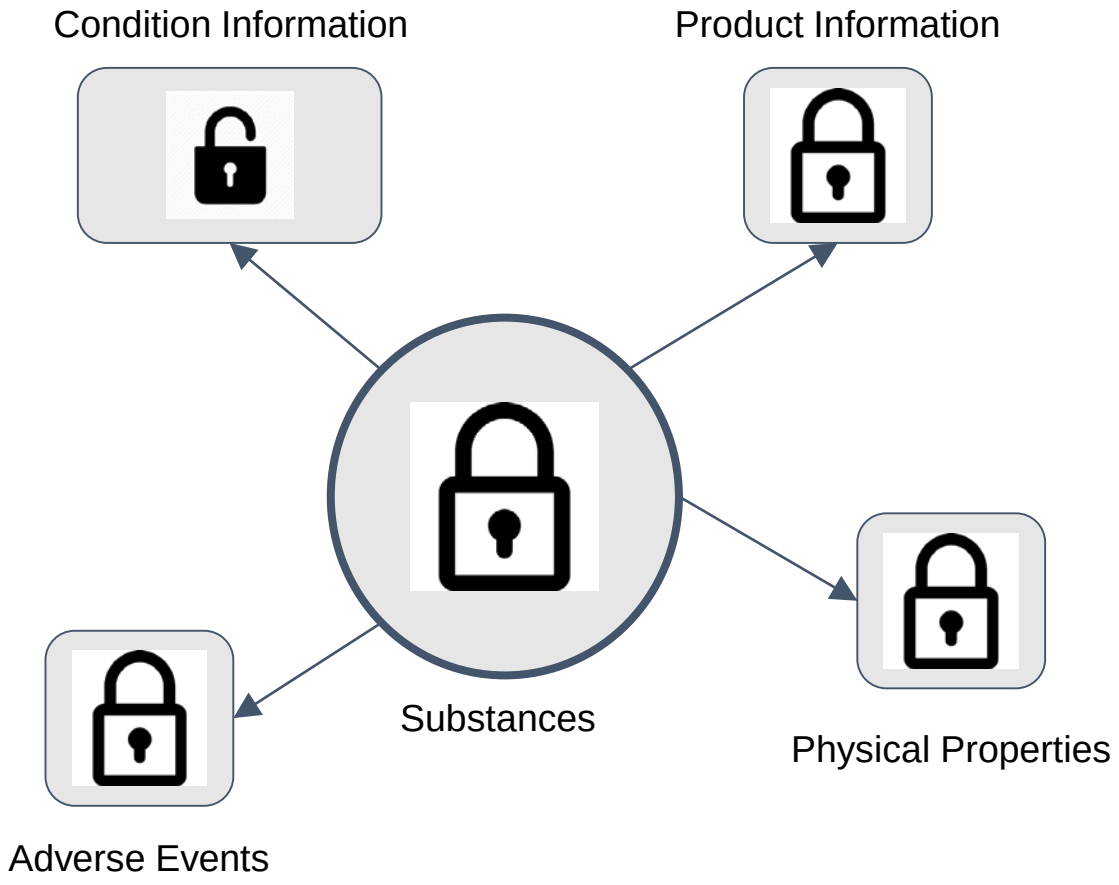
Technical Challenges (cont)

- The main purpose of GSRS is to represent substances faithfully and uniquely
- But the *reason* we want that is to make it easier to link data to those substances



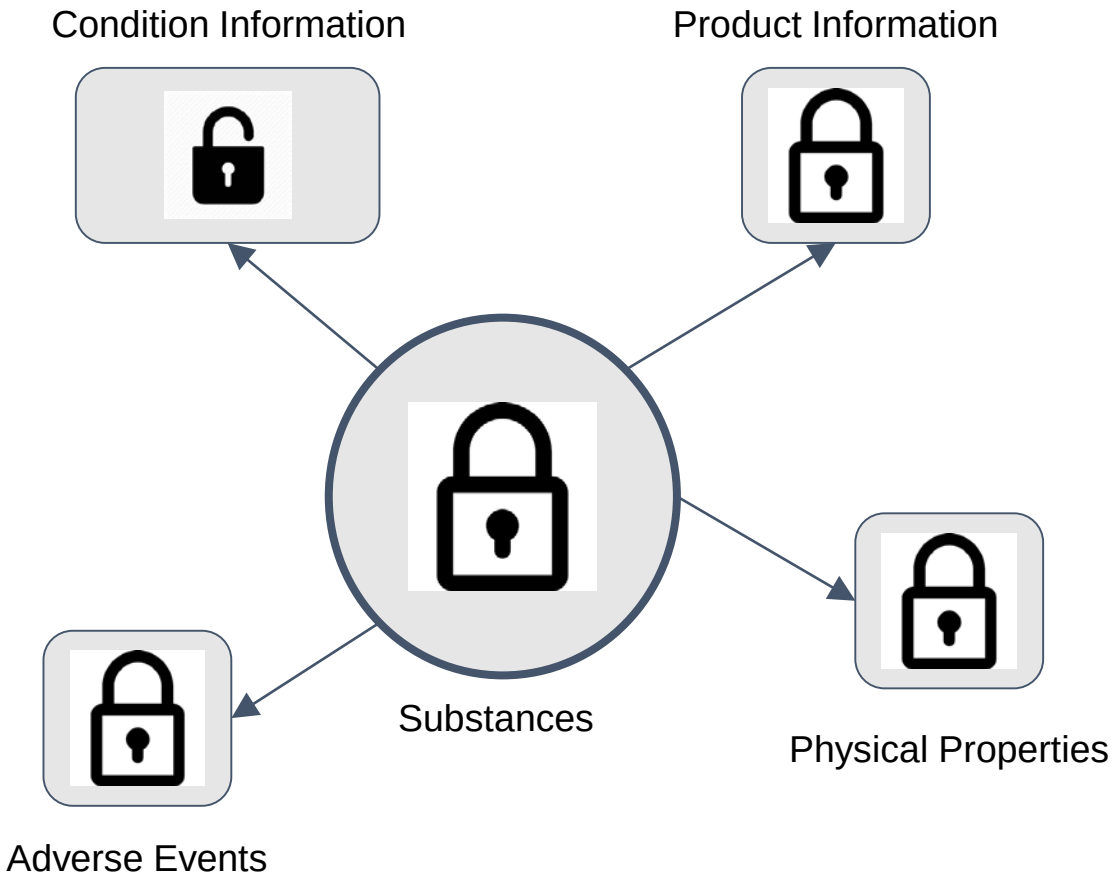
Technical Challenges (cont)

- FDA GSRS is a balance between developing rigorous models, and demonstrating utility
- GSRS needed to become more flexible in order to link some under-standardardized data to substances



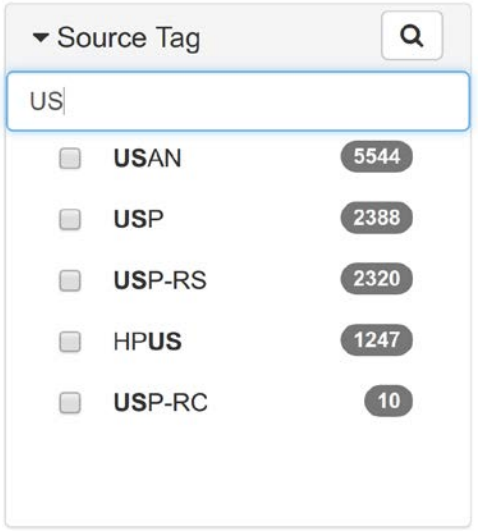
Technical Challenges (cont)

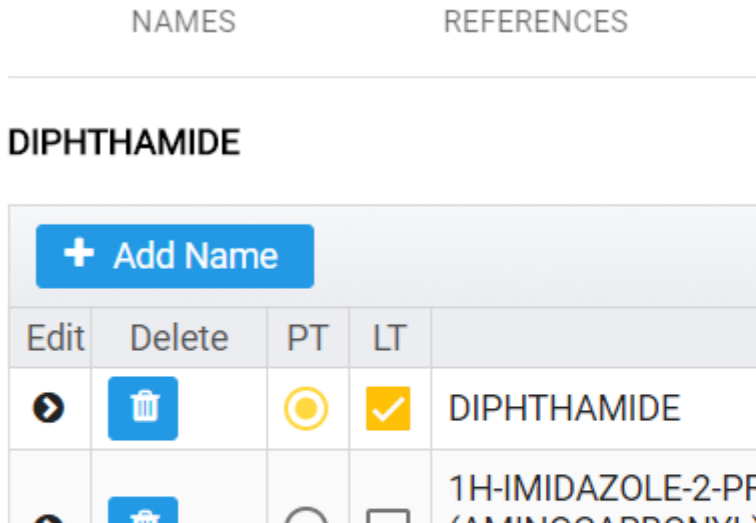
- Well-defined and locked-down substance information can be communicated via a single JSON message
- However, other, free-form information also exists, and must be captured until it can be standardized.



Short List of new Software Features in 2.0

- **Improved Facet Support**
  - Searching for values inside a filter
  - Easy to integrate / configure new facets/filters
- **New Simple Forms for quick updates**
- **New Export Model**
  - Store results to profile for later reference
  - Rerun stored exports and schedule repeated runs
  - Easy interface for developing new exports / transforms
- **Improved Flexibility in REST API**
- **Scheduler and Job Support**
- **Stability and Scalability Improvements**

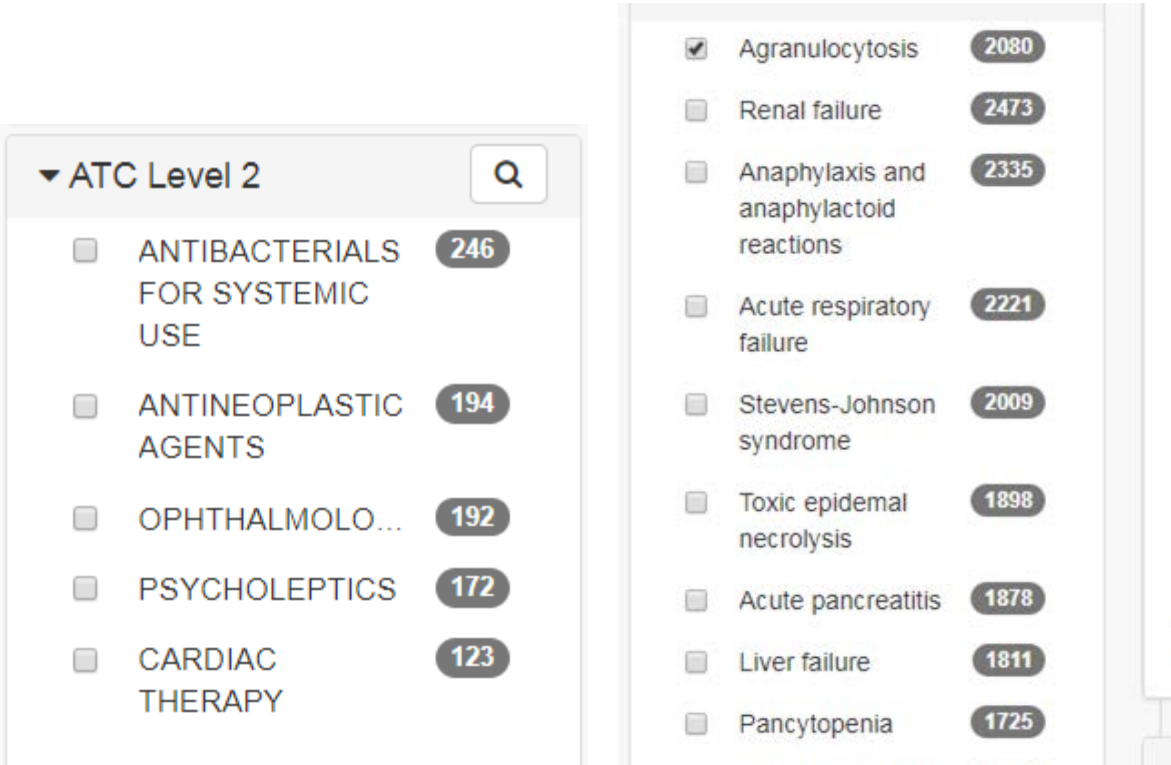


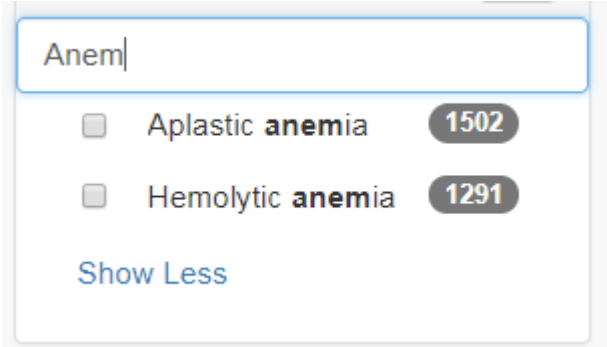


National Center for Advancing Translational Sciences

Improved Facet Support

- **Developers**
  Simple java interfaces which allow new Data or calculations to become Facets
- **Users**
  - Rich data now more easily explored both via UI and REST API





Improved Flexibility in REST API

/names


```
[
 - {
      uuid: "00e4cbad-f33b-43fa-894c-7e5d03ca9360",
      created: 1507659524000,
      createdBy: "admin",
      lastEdited: 1507659524000,
      lastEditedBy: "admin",
      deprecated: false,
      name: "ASPIRIN COMPONENT OF VICOPRIN",
      type: "cn",
      domains: [ ],
    - languages: [
          "en"
      ],
      nameJurisdiction: [ ],
      nameOrgs: [ ],
      preferred: false,
      displayName: false,
    - references: [
          "46e82f2b-8ec5-4d3a-8039-62ba0f7ab34d"
      ],
      access: [ ],
      _self: "https://ginas.ncats.nih.gov/ginas/app/api/v1/names(00e4cbad-f33b-43fa-894c-7e5d03ca9360)?view=full"
  },
 - {
      uuid: "069c37f0-4a19-4c8f-ad34-196556bb5037",
      created: 1507659524000,
      createdBy: "admin",
      lastEdited: 1507659524000,
      lastEditedBy: "admin",
      deprecated: false,
      name: "ASPIRIN",
      type: "cn",
```


Relative paths let you get just the “names”

/names(type:sys)


```
[
 - {
      uuid: "17796a92-7d19-41dd-8a1d-85f7fc2602ac",
      created: 1507659524000,
      createdBy: "admin",
      lastEdited: 1507659524000,
      lastEditedBy: "admin",
      deprecated: false,
      name: "ACETYL SALICYLATE",
      type: "sys",
      domains: [ ],
    - languages: [
          "en"
      ],
      nameJurisdiction: [ ],
      nameOrgs: [ ],
      preferred: false,
      displayName: false,
    - references: [
          "b3548777-047f-44d3-9c55-4909a0fde423"
      ],
      access: [ ],
      _self: "https://ginas.ncats.nih.gov/ginas/app/api/v1/names(17796a92-7d19-41dd-8a1d-85f7fc2602ac)?view=full"
  },
 - {
      uuid: "19144a23-5c43-4ad2-b894-761915330349",
      created: 1507659524000,
      createdBy: "admin",
      lastEdited: 1507659524000,
      lastEditedBy: "admin",
      deprecated: false,
      name: "2-ACETYLOXYBENZOIC ACID",
      type: "sys",
      domains: [ ],
    - languages: [
          "en"
      ],
```


But you can also get only the “Systematic” names


```
      "d87b9521-c829-4f0c-8f66-101f6f3a67b5",
      "fa052f29-7e10-4aab-a1b7-abeaaf7d7b12",
      "81938fb2-747b-471f-a2a4-f82a1e2bf089",
      "6b443371-95be-47a4-901e-9c314a8fb5db"
```

Improved Flexibility in REST API

/names


```
[
  - {
      uuid: "00e4cbad-f33b-43fa-894c-7e5d03ca9360",
      created: 1507659524000,
      createdBy: "admin",
      lastEdited: 1507659524000,
      lastEditedBy: "admin",
      deprecated: false,
      name: "ASPIRIN COMPONENT OF VICOPRIN",
      type: "cn",
      domains: [ ],
    - languages: [
        "en"
      ],
      nameJurisdiction: [ ],
      nameOrgs: [ ],
      preferred: false,
      displayName: false,
    - references: [
        "46e82f2b-8ec5-4d3a-8039-62ba0f7ab34d"
      ],
      access: [ ],
      _self: "https://ginas.ncats.nih.gov/ginas/app/api/v1/names(00e4cbad-f33b-43fa-894c-7e5d03ca9360)?view=full"
    },
  - {
      uuid: "069c37f0-4a19-4c8f-ad34-196556bb5037",
      created: 1507659524000,
      createdBy: "admin",
      lastEdited: 1507659524000,
      lastEditedBy: "admin",
      deprecated: false,
      name: "ASPIRIN",
      type: "cn",
      domains: [ ],
    - languages: [
        "en"
      ],
      nameJurisdiction: [ ],
      nameOrgs: [ ],
      preferred: true,
      displayName: true,
    - references: [
        "a2bacf49-c3ab-42b1-948e-4caf9d13dcc6",
        "c2c31c52-cf50-477b-b020-aa85b4cfc4d4",
        "d87b9521-c829-4f0c-8f66-101f6f3a67b5",
        "fa052f29-7e10-4aab-a1b7-abeaaf7d7b12",
        "81938fb2-747b-471f-a2a4-f82a1e2bf089",
        "6b443371-95be-47a4-901e-9c314a8fb5db"
```


/names(type:sys)!(name)


```
[
    "ACETYL SALICYLATE",
    "2-ACETYLOXYBENZOIC ACID",
    "SALICYLIC ACID ACETATE",
    "2-(ACETYLOXY)BENZOIC ACID"
]
```


.. or just the name parts

Improved Flexibility in REST API

/names


```
[
  - {
      uuid: "00e4cbad-f33b-43fa-894c-7e5d03ca9360",
      created: 1507659524000,
      createdBy: "admin",
      lastEdited: 1507659524000,
      lastEditedBy: "admin",
      deprecated: false,
      name: "ASPIRIN COMPONENT OF VICOPRIN",
      type: "cn",
      domains: [ ],
    - languages: [
        "en"
      ],
      nameJurisdiction: [ ],
      nameOrgs: [ ],
      preferred: false,
      displayName: false,
    - references: [
        "46e82f2b-8ec5-4d3a-8039-62ba0f7ab34d"
      ],
      access: [ ],
      _self: "https://ginas.ncats.nih.gov/ginas/app/api/v1/names(00e4cbad-f33b-43fa-894c-7e5d03ca9360)?view=full"
    },
  - {
      uuid: "069c37f0-4a19-4c8f-ad34-196556bb5037",
      created: 1507659524000,
      createdBy: "admin",
      lastEdited: 1507659524000,
      lastEditedBy: "admin",
      deprecated: false,
      name: "ASPIRIN",
      type: "cn",
      domains: [ ],
    - languages: [
        "en"
      ],
      nameJurisdiction: [ ],
      nameOrgs: [ ],
      preferred: true,
      displayName: true,
    - references: [
        "a2bacf49-c3ab-42b1-948e-4caf9d13dcc6",
        "c2c31c52-cf50-477b-b020-aa85b4cfc4d4",
        "d87b9521-c829-4f0c-8f66-101f6f3a67b5",
        "fa052f29-7e10-4aab-a1b7-abeaaf7d7b12",
        "81938fb2-747b-471f-a2a4-f82a1e2bf089",
        "6b443371-95be-47a4-901e-9c314a8fb5db"
```


/names(type:sys)!(name)!limit(1)

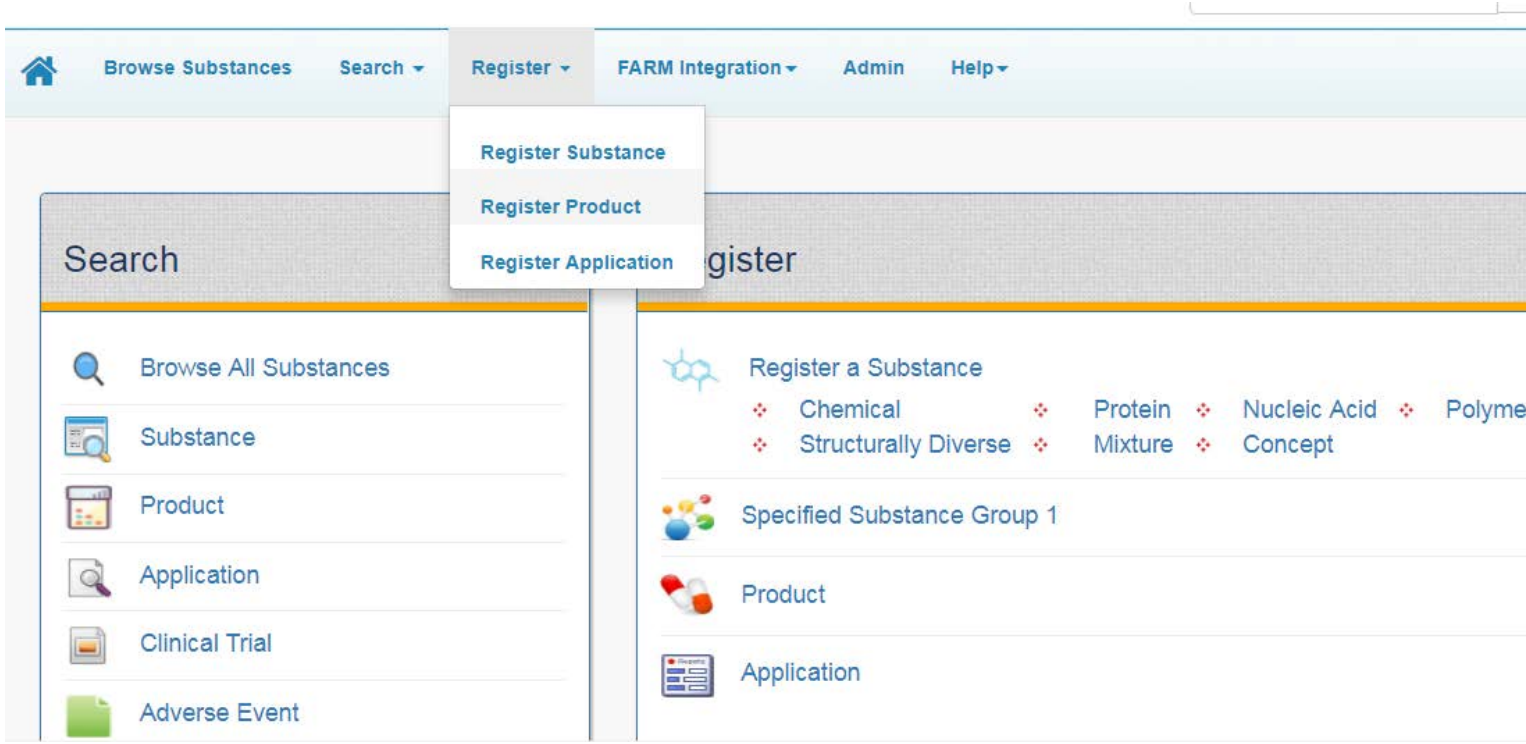

```
[
    "ACETYL SALICYLATE"
]
```


.. or just the name parts

DEMO

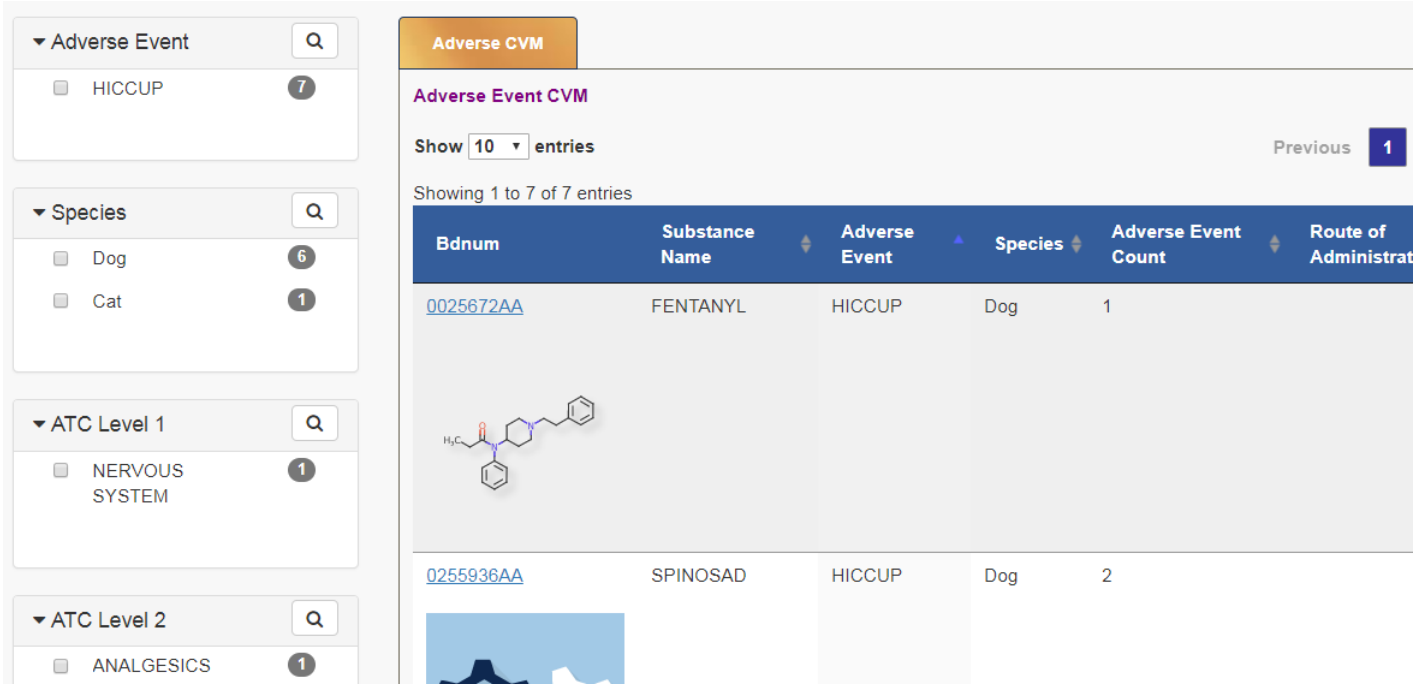
FDA-Specific Enhancements

- **Additional Registration**
  - Products
  - Applications



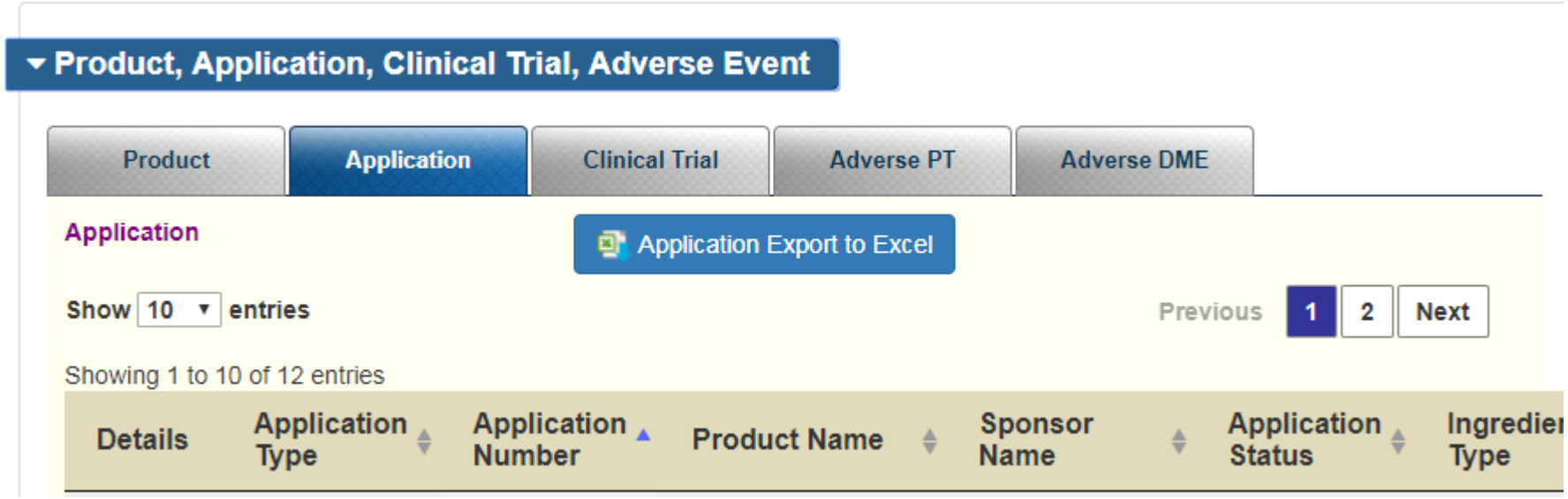
FDA-Specific Enhancements

- Additional Registration
- **Advanced Search**
  - Adverse Events
  - Clinical Trials
  - Products
  - Applications



FDA-Specific Enhancements

- Additional Registration
- Advanced Search
- **Dashboard View of Substance**



Next Steps

- First quarter next year
  - GSRS 2.1
    - Minor additions for consistency with new ISO standard
    - Basic Group 2 and 3 Specified Substances
    - “Additional Data” finders
    - Large Nucleic Acid support
    - Improvements to document handling
  - Standardized pipeline for public data updates
  - Merging additions into single package

Next Next steps

- 2018
  - Basic Helm format transformation support
  - Docker Image
  - Packaged suite of GSRS tools
  - Detailed developer’s guide for extending GSRS software

Thank You