DATA WAREHOUSE
STUDENT DATA USER GROUP
NOVEMBER 13, 2019

Agenda

New Student data warehouse collection

___

- Timeline - Design, Development, Testing

- Timing for Production data – into NGSS, in new DW tables

- Transition planning

- Deep Dive – new tables for Course data
  - COURSE, COURSE _GRADE_MODE, COURSE_XLIST
  - CRSE_SECTION, CRSE_SECT_SCHED_WITH_GRP, CRSE_SECT_SCHED_WITH

- News items

Student Data Warehouse Timeline

Design

- New table structures definition complete

- Initial set of corporate documents defined

- WebI universe design will start November 2019, scheduled to complete February 2020.

- Technical specifications approximately 75% complete

Student Data Warehouse Timeline

Development

- Development started August 2019

- Estimated completion of table build and ETL processes January 2020

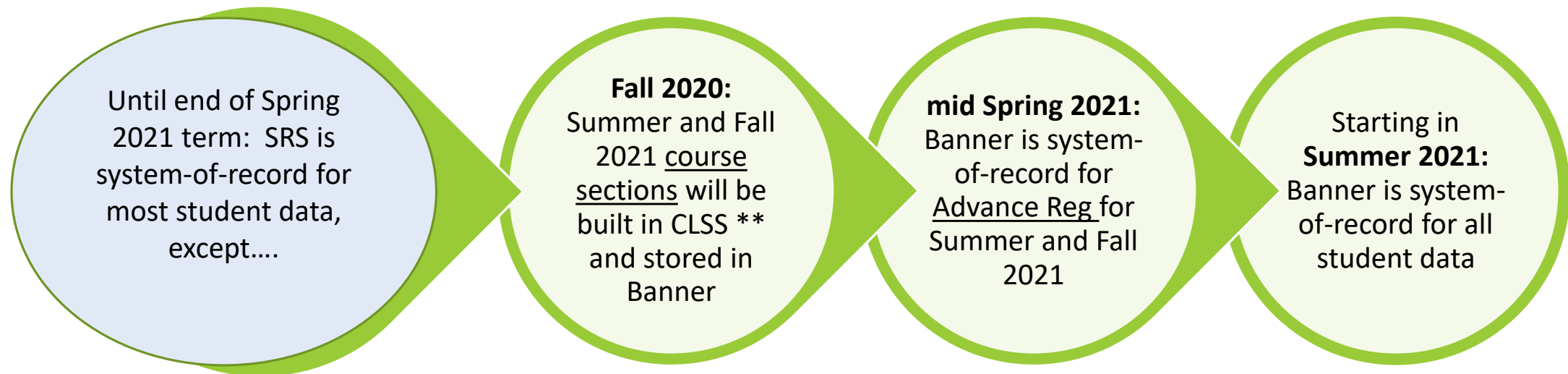- WebI universe development complete March 2020

Student Data Warehouse Timeline

Testing

- Initial test phase for internal project team and early adopters scheduled for October 2019-March 2020

- Availability of tables and WebI universe to wider Penn Community in April 2020

- Initial data set will be limited test data available in the Functional Object Testing instance of Banner

- Test data of total converted population available in June 2020

- Initial set of Corporate Documents available in June 2020

Timing for production data *: when will "my data" start going into NGSS?

Until end of Spring 2021 term: SRS is system-of-record for most student data, except….

Fall 2020: Summer and Fall 2021 <u>course sections</u> will be built in CLSS ** and stored in Banner

mid Spring 2021: Banner is system-of-record for <u>Advance Reg</u> for Summer and Fall 2021

Starting in **Summer 2021:** Banner is system-of-record for all student data

* Test data will be available sooner… this slide speaks to the actual production data

** CLSS is the 3rd-party system in which departmental admins will maintain course sections

Timing for production data: when will "my data" be available in the new warehouse tables?

- **Fall term 2020**: if you are looking for **course sections that will be taught in Summer or Fall 2021**, look in the new DWNGSS_PS tables. Else, look in the old data collection to report on courses and sections taught in Spring 2021 or earlier.

- **Mid-Spring term 2021**: if you are looking for **registration in Summer or Fall 2021** courses, look in the new DWNGSS_PS tables. Else, look in the old data collection. NOTE: reporting on Spring 2021 graduation should be done out of the old data collection.

- **Starting in Summer term 2021**:
  - All students from SRS will have their BIO data converted into Banner.
  - All students who were active in 2010 or any time since then will have all of their Academic activity converted into Banner.
  - There will be a period of time in which you will still need to go back to the old legacy data collection for students who have not been active since 2010.

Timing: to help us plan for this transition, we need your input

- What reports do you typically run in May/June/July every academic year?

- Of those, what reports are the most critical?

- Please take a look at the proposed list of canned reports (see next slide) and let us know if you have a critical report that isn't covered in some fashion by that list.
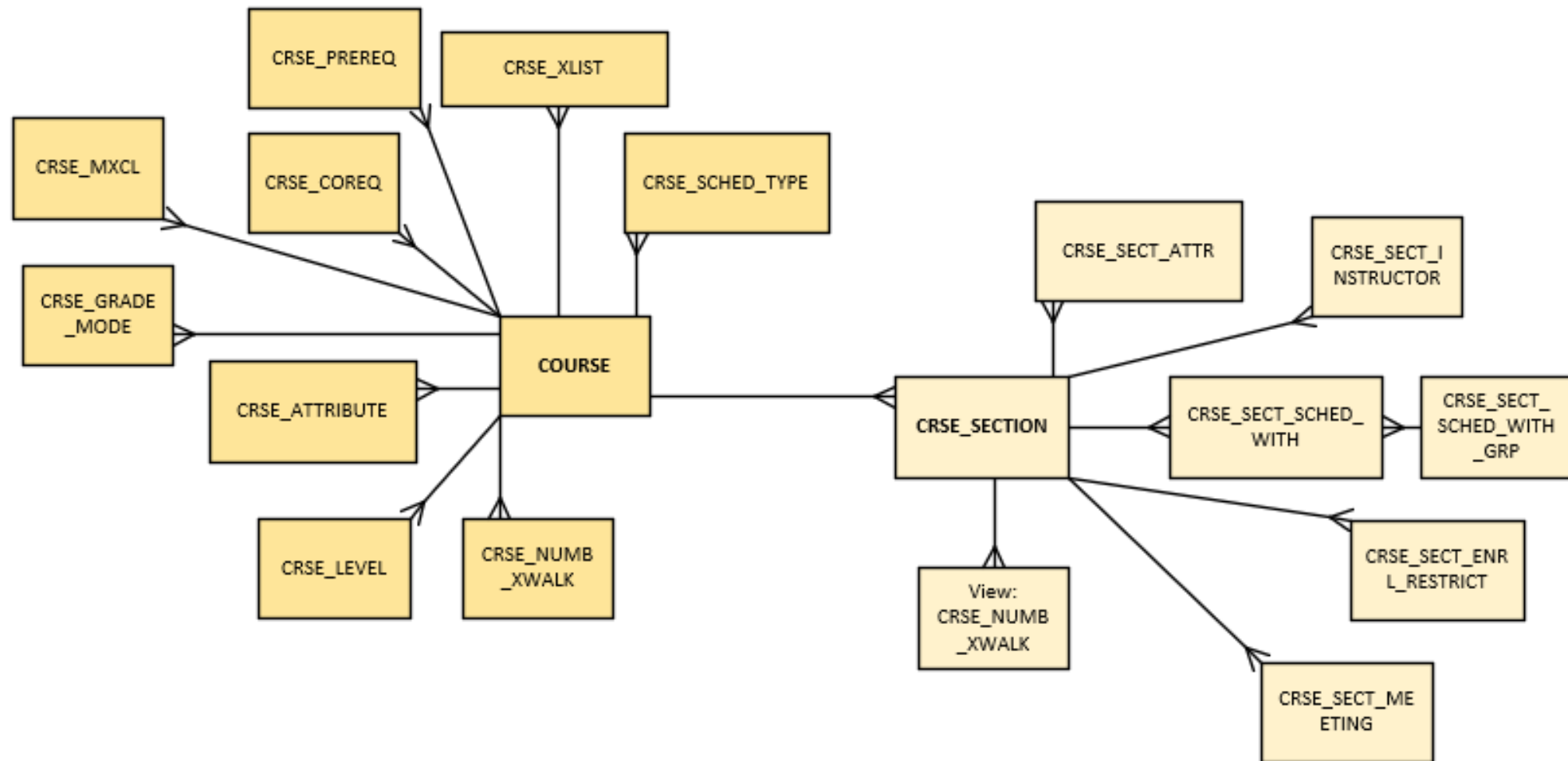
Proposed Pennant Records corporate documents:
https://upenn.box.com/s/7bwnyvhwgg4b1tsgccjtm9fnw5terqw9

Short Description	Category	Business Purpose	Potential user-supplied parameters	Replicates existinging report?	Fre of
Count Students with no Tuition Billed	Cross-collection	Find students who are actively enrolled in a term but who were not billed tuition	Banner college, degree, major, type of student, type of program, term	yes	var
Count PhD Students and Stipends	Cross-collection, F	Find PhD students who have stipend funding	Grad Group, term	yes	var
Students on Hold	Hold	Find students who are currently on hold	Banner college, major, type of student, term	no- new report	var
Student Degree List	Graduation	Be able to report on students who are graduating, or who have graduated in a specified term(s)	Banner college, major, type of student, term	yes	var
Faculty Load	Courses and Instr	Instructors teaching in a term (excluding cancelled courses), with the instructor and course attributes	college and department	sort-of	var
		Report on courses that are "also offered as" under			

The list has a short name and description of the proposed report, what user-supplied parameters might be available, and other properties like the intended audience, etc.

Deeper dive: new tables for COURSE data

COURSE, COURSE _GRADE_MODE, COURSE_XLIST
CRS_SECTION, CRS_SECT_SCHED_WITH_GRP, CRS_SECT_SCHED_WITH

COURSE

COURSE_ID	SUBJECT	COURSE_NUM	EFFECTIVE_TERM	TERM	CATALOG_START_TERM	CATALOG_END_TERM	SCHOOL	DEPARTMENT	ACTIVE_FLAG	TITLE	LONG_TITLE	COURSE_ADMIN_CODE
AFRC0002	AFRC	0002	202120	202120	202120	999999	A	AFRC	A	Intro to Sociology	Introduction to Sociology	2C
AFRC0002	AFRC	0002	202120	202130	202120	999999	A	AFRC	A	Intro to Sociology	Introduction to Sociology	2C
AFRC0002	AFRC	0002	202210	202210	202120	999999	A	AFRC	A	Intro to Sociology	Introduction to Sociology and Culture	2C
AFRC0002	AFRC	0002	202210	202220	202120	999999	A	AFRC	A	Intro to Sociology	Introduction to Sociology and Culture	2C
SOCI0001	SOCI	0001	202120	202120	202120	999999	A	SOCI	A	Intro to Sociology	Introduction to Sociology	2C
SOCI0001	SOCI	0001	202120	202130	202120	999999	A	SOCI	A	Intro to Sociology	Introduction to Sociology	2C
SOCI0001	SOCI	0001	202210	202210	202120	999999	A	SOCI	A	Intro to Sociology	Introduction to Sociology and Culture	2C
SOCI0001	SOCI	0001	202210	202220	202120	999999	A	SOCI	A	Intro to Sociology	Introduction to Sociology and Culture	2C

The COURSE table (i.e., course inventory) will store records term-by-term. The following slides explain why…

COURSE, if we used term-effective

SUBJECT	COURSE_NUM	EFFECTIVE_TERM	SCHOOL	DEPARTMENT	LONG_TITLE
AFRC	0002	202120 (Summer 2021)	A	AFRC	Introduction to Sociology
AFRC	0002	202210 (Spring 2022)	A	AFRC	Introduction to Sociology and Culture

In the source database, Course has a "term effective" row

CRSE_GRADE_MODE, term-effective

SUBJECT	COURSE_NUM	EFFECTIVE_TERM	GRADE_MODE_CODE
AFRC	0002	202120 (Summer 2021)	S (standard grade)
AFRC	0002	202130 (Fall 2021)	P (pass/fail)
AFRC	0002	202130 (Fall 2021)	S (standard grade)

Each of the tables that are related to COURSE, for example here the Grade Mode table, also is stored in the source using term-effective. They each can have different term-effective terms, and can be different from the term-effective term on the base COURSE table. This makes joining between tables difficult.

COURSE term-by-term

SUBJECT	COURSE_NUM	TERM	SCHOOL	DEPARTMENT	LONG_TITLE		EFFECTIVE_TERM
AFRC	0002	202120 (Summer 2021)	A	AFRC	Introduction to Sociology	Everything about these first two records is exactly the same, except the TERM. We will have a record for every possible TERM that this definition is in effect.	202120 (Summer 2021)
AFRC	0002	202130 (Fall 2021)	A	AFRC	Introduction to Sociology		202120 (Summer 2021)
AFRC	0002	202210 (Spring 2022)	A	AFRC	Introduction to Sociology and Culture	As soon as something changes, we get a new Effective Term.

From then on everything is the same except the TERM. We will have a record for every possible TERM that this definition is in effect. | 202210 (Spring 2022) |
| AFRC | 0002 | 202220 (Summer 2022) | A | AFRC | Introduction to Sociology and Culture | | 202210 (Spring 2022) |
| AFRC | 0002 | 202230 (Fall 2022) | A | AFRC | Introduction to Sociology and Culture | | 202210 (Spring 2022) |

So instead, in the warehouse, we will store all Course related tables with one row per course per term, making it easy to join together everything that was happening for any given course in any given term.

CRSE_GRADE_MODE term-by-term

SUBJECT	COURSE_NUM	TERM	GRADE_MODE_CODE	EFFECTIVE_TERM
AFRC	0002	202120 (Summer 2021)	S (standard grade)	202120 (Summer 2021)
AFRC	0002	202130 (Fall 2021)	P (pass/fail)	202130 (Fall 2021)
AFRC	0002	202130 (Fall 2021)	S (standard grade)	202130 (Fall 2021)
AFRC	0002	202210 (Spring 2021)	P (pass/fail)	202130 (Fall 2021)
AFRC	0002	202210 (Spring 2021)	S (standard grade)	202130 (Fall 2021)

The tables that are all related to COURSE can have multiple rows per term. For example, in the Grade Mode table, you will see one row for each allowable grade mode for each course in every term.

CRSE_XLIST term-by-term

COURSE_ID	SUBJECT	COURSE_NUM	TERM	XLIST_FAMILY	PRI_SEC_IND	XLIST_PRI_SUBJECT	XLIST_PRI_COURSE_NUM	EFFECTIVE_TERM	ACTIVITY_DATE
AFRC002	AFRC	0002	202120	7140	P	AFRC	0002	202120	11/8/2019 14:48
SOCI0001	SOCI	0001	202120	7140	S	AFRC	0002	202120	11/8/2019 14:48
AFRC002	AFRC	0002	202130	7140	P	AFRC	0002	202120	11/8/2019 14:48
SOCI0001	SOCI	0001	202130	7140	S	AFRC			
AFRC002	AFRC	0002	202210	7140	P	AFRC			
SOCI0001	SOCI	0001	202210	7140	S	AFRC			
AFRC002	AFRC	0002	202220	7140	P	AFRC			
SOCI0001	SOCI	0001	202220	7140	S	AFRC			
AFRC002	AFRC	0002	202230	7140	P	AFRC			
SOCI0001	SOCI	0001	202230	7140	S	AFRC			
ANTH0003	ANTH	0003	202230	7140	S	AFRC	0002	202230	11/8/2019 14:48

An example of what the course Crosslist table will look like.
Notice that this table will tell you:
- The crosslist "family" id
- Whether the course is the Primary, or one of the Secondary members of the family
- References back to the primary from each of the secondaries.

CRSE_SECTION

TERM	SECTION_ID	SUBJECT	COURSE_NUM	SECTION_NUM	SCHOOL	DIVISION	DEPARTMENT	PART_OF_TERM	PTRM_START_DATE	PTRM_END_DATE	PTRM_WEEKS	SECTION_STATUS	SCHEDULE_TYPE	CAMPUS	TITLE
202210	AFRC0002001	AFRC	0002	001	A	AU	AFRC	1	1/12/2022	5/10/2022	16	A	LEC	PHL	Intro to Sociology
202210	AFRC0002011	AFRC	0002	011	A	AU	AFRC	1	1/12/2022	5/10/2022	16	A	REC	PHL	Intro to Sociology
202210	AFRC0002012	AFRC	0002	012	A	AU	AFRC	1	1/12/2022	5/10/2022	16	A	SEM	PHL	Intro to Sociology
202210	SOCI0001001	SOCI	0001	001	A	AU	SOCI	1	1/12/2022	5/10/2022	16	A	LEC	PHL	Intro to Sociology
202210	SOCI0001011	SOCI	0001	011	A	AU	SOCI	1	1/12/2022	5/10/2022	16	A	REC	PHL	Intro to Sociology
202210	SOCI0001012	SOCI	0001	012	A	AU	SOCI	1	1/12/2022	5/10/2022	16	A	SEM	PHL	Intro to Sociology

The Course Section table has one row per scheduled section offering of a course (each schedule type is a separate section) per term.

CRSE_SCHED_WITH (for an xlist family)

TERM	SCHED_WITH_GROUP	COURSE_REF_NUM	SECTION_ID	SUBJECT	COURSE_NUM	SECTION_NUM	XLIST_FAMILY	PRI_SEC_IND
202210	9786	34567	AFRC0002001	AFRC	0002	001	7140	P
202210	657	54387	AFRC0002011	AFRC	0002	011	7140	P
202210	658	14790	AFRC0002012	AFRC	0002	012	7140	P
202210	9786	5432	SOCI0001001	SOCI	0001	001	7140	S
202210	657	87432	SOCI0001011	SOCI	0001	011	7140	S
202210	658	9	SOCI0001012	SOCI	0001	012	7140	S

Course sections that are offered at the same time in the same room taught by the same instructors are "Scheduled With". Cross-listed courses are just one type of "scheduled with" sections. Sections that are "scheduled with" each other are assigned to the same "**scheduled with group**"

CRSE_SCHED_WITH_GRP

TERM	SCHED_WITH_GROUP	SCHED_WITH_GROUP_DESC	GROUP_MAX_ENRLMNT	GROUP_ENROLLMENT	GROUP_SEATS_AVAIL
202210	9786	Intro Sociology cross-lists Lec	30		
202210	2435	Modern Architecture	25		

The "scheduled with group" controls the number of students that can enroll.

CRSE_SCHED_WITH (non xlist example)

TERM	SCHED_WITH_GROUP	COURSE_REF_NUM	SECTION_ID	SUBJECT	COURSE_NUM	SECTION_NUM	XLIST_FAMILY	PRI_SEC_IND	SECT_ACTIVITY_DATE
202210	2435	456709	ARTH0271001	ARTH	0271	001			########
202210	2435	243761	ARTH6710001	ARTH	6710	001			########

Not all "scheduled with" sections are cross-listed. Here is an example of a graduate course and an undergraduate course that are "scheduled with" each other, but they are not cross-listed.

COURSE data: key points to remember

- COURSE is now connected to a cluster of tables with information related to the definition of a course – not everything about the course inventory is found in just one table anymore.

- Even though the Banner source tables will be "term-effective" we will store COURSE data with a row for each term that course could potentially be used by a section record in CRSE_SECTION.

- Section information -- CRSE_SECTION – is also connected to a cluster of tables with information about all the sections being offered in a specific term. These can be joined to each other on the course section identifier and term. These can also be joined back to the related COURSE record on Subject, Course Number, and Term.

COURSE data: key points to remember

- Cross-listing is defined in COURSE and used in the "scheduled with" relationships in CRSE_SECT*

- "Scheduled with" are courses taught at the same day & time, in the same room, by the same instructor. There can be things that are "scheduled with" that are not cross-listed.

- You will be able to report on BOTH "scheduled with" and cross-listed courses. They can be identified separately, and they can be aggregated together, depending on your needs.

- Credit currency (CU/SH/CR) is defined in COURSE and used in CRSE_SECT* tables.

News

- Pennkey was added to PERSON_ALL_V. No longer need to select from PCOM tables.

- Reminder: It is possible for students added to SRS to *not* have a Penn_ID for 24 hours. Downstream processes should account for this scenario.

- Naming of NWCH New College House to **Lauder College House**
  - New record LDCH "Lauder College House" added to SRS tables
  - Both codes/names remain in dwadmin.BUILDING, COLLEGE_HOUSE tables.

- Naming of PennLaw to **Carey Law School**
  - Not done in SRS yet
  - Change will update school_name in dwadmin.SCHOOL